NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

DISSERTATION

A DISCOURSE ON HUMAN SYSTEMS INTEGRATION

by

Anthony P. Tvaryanas

September 2010

Dissertation Supervisor: Nita Lewis Shattuck

This thesis was done at the MOVES Institute.
Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** September 2010	**3. REPORT TYPE AND DATES COVERED** Dissertation

4. TITLE AND SUBTITLE: A Discourse on Human Systems Integration	**5. FUNDING NUMBERS**
6. AUTHOR(S) Anthony P. Tvaryanas	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A	**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. Protocol No. NPS20090098-IR-EP7-A.

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

This dissertation tackles, head on, two fundamental questions: *What is human systems integration (HSI) and how should one think about HSI problems?* The objective was to develop a coherent systems method to improve the integration of HSI domains to create sustainable systems while preserving system stakeholder preferences.

This dissertation addresses these questions by accomplishing two things: 1) extracting the lessons learned from a historical analysis of the emergence of HSI both as a philosophy and as a Defense Department program, and 2) using those lessons to characterize and illustrate a technical approach to addressing HSI considerations early in an acquisition process. It is shown that the discourse on general systems that occurred over the latter half of the twentieth century, coupled with pressing organizational factors within the U.S. Army, were the principal forces that shaped and drove the emergence and formal recognition of HSI. As determined from this historical analysis, HSI involves the integration of the behavioral sciences, human factors engineering, and operations research to more broadly represent human considerations in early weapon system analyses and the products that evolve from these analyses.

Inclusion of HSI in system analyses necessitates a holistic perspective of the performance and economic trade space formed by the synthesis of the HSI domains. As a result, individual domain interventions are considered in terms of tradeoff decisions. Ideally, the HSI trade space can be systematically explored by integrating Simon's research strategy, Kennedy and Jones' isoperformance approach, and coupling isoperformance with utility analysis through means such as physical programming. Although domain tradeoffs are a central element of HSI, very few studies illustrate the integration of the behavioral sciences and human factors engineering with the tools and methodologies of operations research. Accordingly, three case studies are presented: a preexisting opportunistic dataset of potential Air Force unmanned aircraft pilots, a prospective dataset of Army Soldiers in Basic Combat Training, and data derived from simulation of staffing and shift scheduling solutions using a biomathematical model. Lastly, guidelines for a New HSI method and future challenges are discussed.

14. SUBJECT TERMS Acquisition, Fatigue(Physiology), Human Factors Engineering, Human Systems Integration, Humanities, Isoperformance, Manpower, Operations Research, Performance(Human), Safety, Selection, Social Sciences, Systems Engineering, Training	**15. NUMBER OF PAGES** 642
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

A DISCOURSE IN HUMAN SYSTEMS INTEGRATION

Anthony P. Tvaryanas
Lieutenant Colonel, United States Air Force
B.S., The George Washington University, 1993
M.D., Uniformed University of the Health Sciences, 1997
M.P.H. & M.T.M., Tulane University, 2002

Submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY IN MODELING, VIRTUAL ENVIRONMENTS, AND SIMULATION (MOVES)

from the

NAVAL POSTGRADUATE SCHOOL
September 2010

Author: ____________________________________
Anthony P. Tvaryanas

Approved by:

Nita Lewis Shattuck, Ph.D. Professor of OR/HSI & MOVES Dissertation Supervisor	Eugene P. Paulo, Ph.D. Professor of SE & MOVES
Lawrence G. Shattuck, Ph.D. Professor of OR/HSI	Clifford A. Whitcomb, Ph.D. Chair and Professor of SE
Michael E. McCauley, Ph.D. Professor of OR/HSI & MOVES	

Approved by: ____________________________________
Mathias Kolsch, Chair, MOVES Academic Committee

Approved by: ____________________________________
Peter Denning, Chair, Computer Science Department

Approved by: ____________________________________
Douglas Moses, Associate Provost for Academic Affairs

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This dissertation tackles, head on, two fundamental questions: *What is human systems integration (HSI) and how should one think about HSI problems?* The objective was to develop a coherent systems method to improve the integration of HSI domains to create sustainable systems while preserving system stakeholder preferences.

This dissertation addresses these questions by accomplishing two things: 1) extracting the lessons learned from a historical analysis of the emergence of HSI both as a philosophy and as a Defense Department program, and 2) using those lessons to characterize and illustrate a technical approach to addressing HSI considerations early in an acquisition process. It is shown that the discourse on general systems that occurred over the latter half of the twentieth century, coupled with pressing organizational factors within the U.S. Army, were the principal forces that shaped and drove the emergence and formal recognition of HSI. As determined from this historical analysis, HSI involves the integration of the behavioral sciences, human factors engineering, and operations research to more broadly represent human considerations in early weapon system analyses and the products that evolve from these analyses.

Inclusion of HSI in system analyses necessitates a holistic perspective of the performance and economic trade space formed by the synthesis of the HSI domains. As a result, individual domain interventions are considered in terms of tradeoff decisions. Ideally, the HSI trade space can be systematically explored by integrating Simon's research strategy, Kennedy and Jones' isoperformance approach, and coupling isoperformance with utility analysis through means such as physical programming. Although domain tradeoffs are a central element of HSI, very few studies illustrate the integration of the behavioral sciences and human factors engineering with the tools and methodologies of operations research. Accordingly, three case studies are presented: a preexisting opportunistic dataset of potential Air Force unmanned aircraft pilots, a prospective dataset of Army Soldiers in Basic Combat Training, and data derived from simulation of staffing and shift scheduling solutions using a biomathematical model. Lastly, guidelines for a New HSI method and future challenges are discussed.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I. INTRODUCTION 1
- A. ORIENTATION 1
- B. BIBLIOGRAPHY 9

II. HUMAN SYSTEMS INTEGRATION PHILOSOPHY 11
- A. AN INTRODUCTION TO HUMAN SYSTEMS INTEGRATION 11
- B. THE TRANSITION FROM MACHINE AGE TO SYSTEMS AGE 24
- C. GESTALT AND HOLISM 26
- D. EMERGENCE AND HIERARCHY 28
- E. SYSTEMS TYPOLOGIES 31
- F. COMPLEX ADAPTIVE SYSTEMS 35
- G. WICKED PROBLEMS 38
- H. SOFT AND HARD SYSTEMS APPROACHES 41
- I. TOTAL SYSTEMS INTERVENTION 44
- J. UNDERSTANDING HSI THROUGH TOTAL SYSTEMS INTERVENTION 53
  - 1. Introduction 53
  - 2. Background 54
  - 3. Consulting Using Soft Systems Methodology Through TSI 56
    - *a. TSI: Task, Tool, and Outcome* *56*
    - *b. Creativity* *57*
    - *c. Choice* *58*
    - *d. Implementation* *60*
    - *e. Conclusions From the Workshop Example* *70*
- K. BRINGING IT ALL TOGETHER 71
- L. AN HSI HYPOTHESIS 73
- M. FINAL THOUGHTS 79
- N. BIBLIOGRAPHY 79

III. A BRIEF HISTORY OF THE EMERGENCE OF THE DEFENSE DEPARTMENT'S HUMAN SYSTEMS INTEGRATION PROGRAM 83
- A. PHILOSOPHY VS. PROGRAM 83
- B. SOME CONTEXT 83
  - 1. American Post-World War II Political Culture 83
  - 2. Technological Determinism 91
  - 3. The Evolution of the Army of Excellence 92
  - 4. The Military Reform Movement 110
- C. TECHNOLOGY AND ITS EFFECTS ON PERSONNEL 115
- D. THE DEVELOPMENT OF MANPRINT 127
  - 1. Progenitor Ideas (1930s–1950s) 127
  - 2. Failed Initial Attempts (1967–1983) 131
  - 3. A Resurgence of Interest (1983–) 138
  - 4. Human Systems Integration (1987–) 156

E. **CONCLUSION** **165**
F. **BIBLIOGRAPHY** **166**

IV. **THE HUMAN SYSTEMS INTEGRATION TRADE SPACE PROBLEM** **173**
A. **INTRODUCTION** **173**
B. **THE NAVAL POSTGRADUATE SCHOOL HSI PROCESS MODEL** **175**
1. **Traditional Conceptual Models of HSI** **175**
2. **A Process Model of HSI** **179**
C. **HSI AS THE PROCESS OF MANAGING THE SUPPLY OF AND DEMAND FOR MANPOWER, PERSONNEL, AND TRAINING (MPT) RESOURCES** **184**
1. **MPT Supply and Demand Management** **184**
2. **A Model for Interfacing MPT Supply and Demand** **189**
a. ***The Macro MPT Supply and Demand Interface Process*** ***189***
b. ***The Micro MPT Supply and Demand Interface Process*** ***192***
c. ***The MPT Capability Tradeoff Analysis*** ***195***
D. **ISOPERFORMANCE** **196**
1. **Considering Aptitude and Training Tradeoffs** **196**
2. **Simon's Research Strategy** **200**
3. **Isoperformance Curves** **202**
4. **Illustrative Example: The Main Tank Gunner** **205**
E. **ISOPERFORMANCE IN THE DESIGN OF COMPLEX SYSTEMS** **212**
1. **Isoperformance as a System Design Philosophy** **212**
2. **Isoperformance Problem Formulation** **214**
a. ***Notation*** ***214***
b. ***Step 1: Find the Performance Invariant Set*** ***215***
c. ***Step 2: Find the Efficient Subset*** ***216***
d. ***Step 3: Select the Final Design*** ***217***
3. **Deterministic Versus Empirical Isoperformance Models** **218**
F. **COUPLING ISOPERFORMANCE WITH UTILITY ANALYSIS** **219**
1. **Problem Statement** **219**
2. **Physical Program Problem Formulation** **221**
a. ***Notation and Model Specifications*** ***221***
b. ***Physical Programming Weight Algorithm*** ***226***
c. ***Physical Programming Problem Model*** ***227***
3. **Example** **228**
G. **BRINGING IT ALL TOGETHER** **230**
1. **Illustrative Example: Enhancing Airline Passenger Screening** **230**
2. **Problem Definition** **231**
3. **Defining the HSI Solution Space** **233**
4. **Evaluating Solutions** **234**
5. **Defining a Tradeoff Function** **239**
6. **Developing Isoperformance Curves** **242**
7. **Designing a System Solution** **247**
H. **CONCLUSION** **252**
I. **BIBLIOGRAPHY** **253**

V. ISORELIABILITY MODELS FOR HUMAN SYSTEMS INTEGRATION DOMAIN TRADEOFFS – CHOOSING A PERSONNEL SUPPLY SOURCE FOR FUTURE UNMANNED AIRCRAFT SYSTEM OPERATORS 259
- **A. INTRODUCTION.........259**
  - **1. Statement of the Problem.........259**
  - **2. Purpose of This Study.........262**
  - **3. Theoretical Perspective.........262**
  - **4. Research Questions and Study Hypotheses.........283**
  - **5. Definition of Terms.........284**
  - **6. Delimitations and Limitations.........284**
- **B. METHODS.........285**
  - **1. Participants.........285**
  - **2. Research Design.........287**
  - **3. Instruments.........287**
  - **4. Procedures.........292**
  - **5. Data Analysis Procedures.........293**
- **C. RESULTS.........298**
  - **1. Basic Maneuvering Task.........298**
  - **2. Landing Task.........305**
  - **3. Reconnaissance Task.........312**
- **D. AGGREGATE ISORELIABILITY ANALYSIS.........321**
  - **1. Mathematical Programming Formulation.........321**
  - **2. Example.........326**
- **E. DISCUSSION.........331**
- **F. APPENDIX.........338**
  - **1. Basic Maneuvering Task Regression Model.........338**
  - **2. Landing Task Regression Model.........347**
  - **3. Reconnaissance Task Regression Model.........358**
- **G. BIBLIOGRAPHY.........367**

VI. HUMAN SYSTEMS INTEGRATION DOMAIN TRADEOFFS IN NON-TECHNICAL SYSTEMS – IMPROVING SOLDIER BASIC COMBAT TRAINING.........371
- **A. INTRODUCTION.........371**
- **B. EFFECTS OF SLEEP ON TRAINING EFFECTIVENESS IN SOLDIERS AT FORT LEONARD WOOD.........372**
  - **1. Statement of the Problem.........372**
  - **2. Purpose of the Study.........376**
  - **3. Theoretical Perspective and HSI Model Elaboration.........377**
  - **4. Study Hypotheses.........386**
  - **5. Delimitations and Limitations.........387**
- **C. METHODS.........387**
  - **1. Research Design.........387**
  - **2. Participants.........388**
  - **3. Data Collection Instruments and Variables.........388**
    - ***a. Actiwatch.........388***

b. Basic Rifle Marksmanship *389*
c. General Technical Aptitude *389*
d. Mood State *389*
e. Personality *390*
f. Physical Fitness *390*
g. Resilience *390*
h. Sleep Habits *391*
i. Study Questionnaires *392*
4. Procedures **392**
a. General *392*
b. Actigraphy *393*
c. Statistical Analysis *394*
D. RESULTS **395**
1. Participants (*n* = 392) **395**
2. Actigraphy Subsample **400**
a. Participants (n = 95) *400*
b. Total Sleep Time *406*
c. Sleep Efficiency *412*
d. Activity Counts During Sleep *414*
3. Profile of Mood States **415**
a. Tension-Anxiety (T) Factor *419*
b. Depression-Dejection (D) Factor *420*
c. Anger-Hostility (A) Factor *422*
d. Vigor-Activity (V) Factor *425*
e. Fatigue-Inertia (F) Factor *427*
f. Confusion-Bewilderment (C) Factor *430*
g. Total Mood Disturbance Score *432*
h. Actigraphy Subsample *435*
4. Basic Rifle Marksmanship **444**
5. Physical Fitness **452**
6. Post-Study Questionnaire **470**
a. Epworth Sleepiness Scale *471*
b. Pittsburgh Sleep Quality Index *475*
c. Ordinal Sleep Ratings *478*
d. Frequency of Sleep During Activities *479*
e. Preference in Timing of Physical Fitness Training *480*
7. Attrition **481**
E. DISCUSSION **482**
1. Actigraphic Measures of Sleep **483**
2. Mood States **484**
3. Basic Rifle Marksmanship **487**
4. Physical Fitness **488**
5. Sleep Hygiene **489**
6. Attrition **490**
F. SELECT HUMAN SYSTEMS INTEGRATION ANALYSES **490**

1. Basic Rifle Marksmanship Model 491
2. Sleep Quality Model...497
G. CONCLUSION 500
H. BIBLIOGRAPHY..501

VII. HUMAN SYSTEMS INTEGRATION DOMAIN TRADEOFFS IN OPTIMIZED MANNING – THE TASK EFFECTIVENESS SCHEDULING TOOL 509
A. INTRODUCTION...509
B. PROBLEM STATEMENT 511
C. THE SLEEP, ACTIVITY, FATIGUE, AND TASK EFFECTIVENESS (SAFTE) MODEL 512
D. AN OPERATIONS RESEARCH PERSPECTIVE 514
E. THE BASIC MIXED INTEGER LINEAR PROGRAM.......................515
1. Indices and [Cardinality] 516
2. Data and [Units]...516
3. Variables 517
4. Constraints..517
5. Objective 518
F. RESULTS AND DISCUSSION 519
1. Case 1: High Task Effectiveness Criterion...................................519
2. Case 2: Organizational Hours-of-Work Rules.............................521
3. Case 3: Sleep Quality 523
G. CONCLUSION 524
H. APPENDIX – GAMS CODE 525
I. BIBLIOGRAPHY..528

VIII. LESSONS FROM THE DISCOURSE: WHAT HAVE WE LEARNED?........529
A. OVERVIEW...529
B. SUMMARY OF CONCEPTS..529
C. CASE STUDY REVIEWS 533
D. EPILOGUE 535

IX. A NEW HUMAN SYSTEMS INTEGRATION..539
A. WHAT IS WRONG WITH CLASSIC HUMAN SYSTEMS INTEGRATION?...539
1. Attitudes and Understanding..539
2. Right Approach, Wrong Result?..540
B. A NEW HUMAN SYSTEMS INTEGRATION METHOD....................541
1. Design Guidelines...541
2. Step 1: Establish SOI Objectives and Requirements by Reference to Containing System(s)...542
3. Step 2: Identify Containing System(s)' Strategic Human Resources Objectives 543
4. Step 3: Identify Sibling Systems (Vis-à-vis Shared Human Resources) and Their Interactions That Will Be Perturbed by the SOI 543

5. Step 4: Develop SOI Design Trade Space to Complement Sibling Systems in Contributing to Containing System(s)' Objectives....545
6. Step 5: Functionally Partition SOI and Describe Required (Emergent) Human-System Performance in Terms of Response Surfaces That Are Functions of the Domains of HSI 546
7. Step 6: Reduce Response Surfaces to Isoperformance (Tradeoff) Equations for Incorporation in System Analyses.......547
8. Step 7: Seek a Balanced Design (Joint Optimization) that Satisfices SOI Objectives and Requirements 547
9. Step 8: Continuously Reassess and Rebalance the Design Throughout the Life of the SOI 549

C. CHALLENGES FOR NEW HUMAN SYSTEMS INTEGRATION......549
D. CONCLUSION 552
E. BIBLIOGRAPHY 553

X. APPENDIX: A DISCOURSE IN HUMAN SYSTEMS INTEGRATION BRIEFING SLIDES....555

INITIAL DISTRIBUTION LIST 607

LIST OF FIGURES

Figure I-1. Balls enclosing ball [After Hamming, 1997]. 3
Figure I-2. Causal loop diagram showing the implications of manpower changes. 6
Figure I-3. Intent structure outlining topics and objectives starting with foundation and theory at the bottom and working up to a future vision at the top. Numbered boxes correspond to dissertation chapters. 7
Figure I-4. Bridging from issues to HSI solutions [After Hitchins, 1992]. 8
Figure II-1. HSI double-integration process model [From Booher, 2003]. 12
Figure II-2. Illustration of the Gestalt principle of reification. 27
Figure II-3. Emergence and hierarchy [From Hitchins, 1992]. 28
Figure II-4. Jordan's taxonomy [After Jordan, 1968]. 33
Figure II-5. Checkland's systems map of the universe [From Checkland, 1981]. 34
Figure II-6. Complex adaptive systems. 36
Figure II-7. Spectrum of problems with corresponding systems modeling approaches [From Pidd, 2004]. 39
Figure II-8. Relationships between hard and soft systems approaches [From Pidd, 2004]. 44
Figure II-9. The process of TSI in clockwise and counterclockwise modes [From Flood, 1995]. 48
Figure II-10. Grouping of systems methodologies based upon the assumptions they make about problem contexts [From Flood & Jackson, 1991]. 53
Figure II-11. Presumed relationship between hard and soft systems approaches [After Pidd, 2004]. 54
Figure II-12. The inquiring/learning cycle of soft systems methodology [After Checkland, 2000]. 60
Figure II-13. A representation of HPID's role in the HP Wing (as one of the enabling support systems, S). 61
Figure II-14. Root definition and CATWOE based on Figure II-13. 62
Figure II-15. Rich picture based on Figure II-14. 62
Figure II-16. Concept for systematically developing human performance-related capabilities. 63
Figure II-17. The conceptual model for the root definition in Figure II-14. 64
Figure II-18. Old Human Performance Integration Directorate HSI model. 65
Figure II-19. Concept and root definition for Activity 6. 67
Figure II-20. A rich picture of a system concept for generating human performance solutions. 68
Figure II-21. Activity 6 of Figure II-17 expanded. 68
Figure II-22. Personnel and technological subsystems comprising system-of-interest (SOI) as viewed from that SOI's containing system. 76
Figure II-23 A family of interacting systems, to include a system-of-interest and its sibling systems, all existing within the environment provided by the containing system. 78

Figure III-1. Drawing by the British cartoonist Leslie Illingworth, published in June 1947, depicting the threatening reach of Stalin and the spread of the communist ideology in postwar Europe. 87
Figure III-2. Typical antimilitarism cartoon from 1914 (source unknown). 88
Figure III-3. *Time* magazine cover from April 1957 celebrating two visionary scientist-engineers, Dean Woolridge and Simon Ramo, who are widely credited with introducing the high-technology, science-based, systems-oriented management approach (Hughes 1998)..89
Figure III-4. Drawing by the British cartoonist Leslie Illingworth, published in May 1950, depicting a partitioned globe and emerging communist threats all over the world. 90
Figure III-5. Drawing by Ben Sargent published in September 1981, referencing the budget debate in the U.S. Congress. President Reagan had managed to get major cuts in social security, public service, and welfare prior to that. However, given rising unemployment, people and Congress were increasingly reticent to accept further cuts while the Defense Department was to get their usual funding. Defense is drawn here as an overly obese boxer who is already unable to get up, yet alone fight. The fairly small physique of the general behind him illustrates the fact that the Defense Department leaders had already lost control over the military forces' quantity. The next year, Congress rejected Reagan's plan for more cutbacks and forced a scaling-down of the defense budget (Fischer & Fischer, 1999). 111
Figure III-6. Summary arguments against complexity from the Spinney Report (From Spinney, 1985). 113
Figure III-7. *Time* magazine cover from March 1983 featuring Franklin Spinney, a military analyst for the Pentagon who became famous in the early 1980s for his "Spinney Report" criticizing the perceived reckless pursuit of costly complex weapon systems by the Pentagon. 115
Figure III-8. Growth of technical jobs in the U.S. Army for selected fiscal years 1953–1985 (After Binkin, 1986)..118
Figure III-9. French SOMUA S35s captured by Germany in 1940. 129
Figure III-10. Personnel considerations in system effectiveness (From Department of the Army, 1976)..138
Figure III-11. Depiction of the human considerations addressed in HSI as described in the 1991 edition of DoD Instruction 5000.2 (From DoD, 1991)..................163
Figure IV-1. Forces in equilibrium in simple pulley system [From Hitchins, 1992]. 173
Figure IV-2. Complementary HSI domain inputs neutralizing unwanted perturbations [After Hitchins, 1992]..174
Figure IV-3. Early conceptual HSI model focusing on domain representation [From Department of the Army, 1976]..176
Figure IV-4. Conceptual HSI model focusing on domain representation and the need to consider inter-domain tradeoffs [From Department of Defense, 1991]. 177
Figure IV-5. Systems engineering and HSI processes [From Booher, Beaton, & Greene, 2009]. 178

Figure IV-6. Simplified model of a marine ecosystem [From Miller &Shattuck, 2007]. ..181
Figure IV-7. The Naval Postgraduate School HSI process model with modification by the author [After Miller & Shattuck, 2007]. 182
Figure IV-8. Concept map for MPT supply management as described by DePuy and Bonder (1982). 186
Figure IV-9. Concept map for MPT demand management as described by DePuy and Bonder (1982). 188
Figure IV-10. Schematic representation of the interface between MPT supply and demand at the aggregated level of a military service [After DePuy & Bonder, 1982]. 190
Figure IV-11. Schematic representation of the interface between MPT supply and demand at the system-specific level [After DePuy & Bonder, 1982]). 193
Figure IV-12. Achievable skill levels for subtask X as a function of personnel quality and training level [From DePuy & Bonder, 1982]. 196
Figure IV-13. Possible interactive effects of aptitude and two instructional treatments on learning outcomes [From Whitener, 1989]. 199
Figure IV-14. Equal motion sickness contours [From McCauley & Kennedy, 1976]. 204
Figure IV-15. Function, task hierarchy for a hypothetical combat vehicle system [From DePuy & Bonder, 1982]. 207
Figure IV-16. Surface plot of tracking error as a function of AFQT score (personnel domain of HSI) and training time (training domain of HSI). 209
Figure IV-17. Isoperformance curves trading off aptitude (AFQT score) and training time with the criterion (tracking error) set at 0.60, 0.50, and 0.40 and level of assurance set at 0.50. 211
Figure IV-18. Normalized utility curves u_i for the i^{th} system objective J_i represented by monotonically decreasing (SIB), increasing (LIB) or concave function (NIB) [From de Weck & Jones, 2006]..214
Figure IV-19. Deterministic versus empirical modeling approaches [From de Weck & Jones, 2006]. 218
Figure IV-20. Preference function classification in physical programming [After Messac, 1996]. 222
Figure IV-21. Class function regions for the i^{th} generic criterion [After Messac, Gupta, & Akbulut, 1996]. 223
Figure IV-22. Cartoon by Daryl Cagle from January 14, 2010, commenting on the heightened emphasis on security at the nation's airports following the unsuccessful attack by the underwear bomber [From Cagle, 2010]. 231
Figure IV-23. Three-factor, two-level design. 235
Figure IV-24. Pareto analysis of the airline passenger screening system experiment. 239
Figure IV-25. Potential solution space, shown as shaded areas, for the airline passenger screening system experiment. 242
Figure IV-26. Hypothetical isoperformance plots for two equipment configurations..........243
Figure IV-27. Hypothetical isoperformance plots for two equipment configurations with personnel and training domain feasibility constraints. 243

Figure IV-28. Potential solution space, shown as shaded areas, for the airline passenger screening system experiment with corresponding isoperformance equations. 245
Figure IV-29. Isoperformance curves for the two equipment options: a) Virtually Nude, b) Magic Eyes. 246
Figure V-1. Isoperformance curve trading off aptitude and time on the job, setting the SQT criterion level at 60 and the assurance level at 0.90 [From Jones & Kennedy, 1996]. 267
Figure V-2. Hypothetical isoreliability curves trading off time spent training and personnel category for two distinct and independent tasks performed by a system operator. 273
Figure V-3. Manpower, personnel, and training domains within a system structure [From Hay Systems, 1991, as reproduced in Archer, Headley, & Allender, 2003]. 276
Figure V-4. Effect of human and equipment reliability on overall function reliability [From Nelson, Schmitz, & Promisel, 1984]. 279
Figure V-5. Manpower, personnel, training, and safety domains within a system structure. 283
Figure V-6. Display for the synthetic basic maneuvering task [From Schreiber et al., 2002]. 289
Figure V-7. Primary flight display (left) and tracker map (right) for the synthetic landing task [After Schreiber et al., 2002]. 290
Figure V-8. Primary flight display symbology overlaid on payload camera imagery for the synthetic reconnaissance task [From Schreiber et al., 2002]. 291
Figure V-9. Synthetic task environment equipment setup [From Schreiber et al., 2002]. 292
Figure V-10. Scatter plot of the raw basic maneuvering task data. 299
Figure V-11. Plot of the fitted basic maneuvering task data. 301
Figure V-12. Isoreliability curves trading off aptitude (personnel category) and training time with the proportion proficient set at 0.50, 0.75, 0.90, 0.95, and 0.99 and level of assurance set at 0.90 for all criterion settings. 304
Figure V-13. Personnel sensitivity, expressed in terms of the percent difference in training time for each personnel category relative to the Predator selectee category, for various settings of the reliability criterion on the basic maneuvering task. 305
Figure V-14. Scatter plot of the landing task data. 306
Figure V-15. Plot of the fitted landing task data. 308
Figure V-16. Isoreliability curves trading off aptitude (personnel category) and training trials with the proportion proficient set at 0.50, 0.75, 0.90, 0.95, and 0.99 and level of assurance set at 0.90 for all criterion settings. 311
Figure V-17. Personnel sensitivity, expressed in terms of the percent difference in training time for each personnel category relative to the Predator selectee category, for various settings of the reliability criterion on the landing task. 312
Figure V-18. Scatter plot of the reconnaissance task data for six Predator pilots. 313

Figure V-19. Scatter plot of the reconnaissance task data for experienced Predator pilots versus the fitted models. 314
Figure V-20. Scatter plot of the reconnaissance task data..315
Figure V-21. Plot of the fitted reconnaissance task data. 318
Figure V-22. Isoreliability curves trading off aptitude (personnel category) and training trials with the proportion proficient set at 0.50, 0.75, 0.90, 0.95, and 0.99 and level of assurance set at 0.90 for all criterion settings. 320
Figure V-23. Personnel sensitivity, expressed in terms of the percent difference in training time for each personnel category relative to the Predator selectee category, for various settings of the reliability criterion on the reconnaissance task. 321
Figure V-24. Reliability block diagram...322
Figure V-25. Cost versus total reliability plot of feasible solutions for the Predator operator selection problem (cost effectiveness objective values given in parentheses)..331
Figure V-26. Plots of the generalized linear model of `Proficient` (*y*) predicted by `Time`, `Group`, and `Time:Group.` 340
Figure V-27. Plots of the generalized linear model of `Proficient` (*y*) predicted by log(`Time)`, `Group`, and log(`Time):Group.` 344
Figure V-28. The partial fits for the generalized additive logistic regression model of `Proficient` (y) with `Time`, `Group`, and `Time:Group` as predictors. ..346
Figure V-29. Plots of the generalized linear model of `Proficient` (*y*) predicted by `Trials`, `Group`, and `Trials:Group.` 349
Figure V-30. Plots of the generalized linear model of `Proficient` (*y*) predicted by `log(Trials)`, `Group`, and `log(Trials):Group.` 354
Figure V-31. Plots of the generalized linear model of `Proficient` (*y*) predicted by `Trials`, `Group`, and `Trials:Group`, using `probit` link function. 355
Figure V-32. The partial fits for the generalized additive logistic regression model of `Proficient` (*y*) with `Trials`, `Group`, and `Trials:Group` as predictors...357
Figure V-33. Plots of the generalized linear model of `Proficient` (*y*) predicted by `Time`, `Group`, and `Trial:Group.` 360
Figure V-34. Plots of the generalized linear model of `Proficient` (*y*) predicted by `log(Trial)`, `Group`, and `log(Trial):Group` 362
Figure V-35. The partial fits for the generalized additive logistic regression model of `Proficient` (*y*) with `Trial`, `Group`, and `Trial:Group` as predictors...364
Figure VI-1. Stock and flow diagram of the SAFTE model...378
Figure VI-2. Manpower, personnel, training, safety, survivability, and habitability domains within a system structure...384
Figure VI-3. Hypothetical interactive effects of aptitude and two training schedules on learning outcomes. 385
Figure VI-4. Histogram of participants' reported usual bed time (PSQI question 1).........399

Figure VI-5. Histogram of participants' reported usual getting up time (PSQI question 3).....399
Figure VI-6. Histogram of participants' reported hours of sleep per night (PSQI question 4).....400
Figure VI-7. Histogram of participants' reported usual bed time (PSQI question 1) in actigraphy subsample.....405
Figure VI-8. Histogram of participants' reported usual getting up time (PSQI question 3) in actigraphy subsample.....405
Figure VI-9. Histogram of participants' reported hours of sleep per night (PSQI question 4) in actigraphy subsample.....406
Figure VI-10. Histograms of total sleep obtained at night for all sleep observations gathered during Basic Combat Training according to treatment condition.....407
Figure VI-11. Estimated marginal means for sleep by week of training (error bars are for 99% confidence intervals).....410
Figure VI-12. Estimated marginal means for sleep by chronotype (error bars are for 99% confidence intervals).....411
Figure VI-13. Estimated marginal means for sleep by treatment condition and week of training (error bars are for 99% confidence intervals).....412
Figure VI-14. Histograms of sleep efficiency for all sleep observations by treatment condition.....413
Figure VI-15. Histograms of mean activity counts for all sleep observations by treatment condition.....415
Figure VI-16. Comparison of estimated marginal means and associated 95% confidence intervals for pre-study POMS factor scores by study group.....416
Figure VI-17. Estimated marginal means for POMS T-factor delta from baseline scores by week of training (error bars are for 99% confidence intervals).....420
Figure VI-18. Estimated marginal means for POMS D-factor delta from baseline scores by week of training (error bars are for 99% confidence intervals).....422
Figure VI-19. Estimated marginal means for POMS A-factor delta from baseline scores by week of training (error bars are for 99% confidence intervals).....424
Figure VI-20. Estimated marginal means for POMS A-factor delta from baseline scores by treatment condition and week of training (error bars are for 99% confidence intervals).....425
Figure VI-21. Estimated marginal means for POMS V-factor delta from baseline scores by chronotype (error bars are for 99% confidence intervals).....427
Figure VI-22. Estimated marginal means for POMS F-factor delta from baseline scores by week of training (error bars are for 99% confidence intervals).....429
Figure VI-23. Estimated marginal means for POMS F-factor delta from baseline scores by treatment condition and week of training (error bars are for 99% confidence intervals).....430
Figure VI-24. Estimated marginal means for POMS C-factor delta from baseline scores by week of training (error bars are for 99% confidence intervals).....432
Figure VI-25. Estimated marginal means for POMS Total Mood Disturbance (TMD) delta from baseline scores by week of training (error bars are for 99% confidence intervals).....434

Figure VI-26. Estimated marginal means for POMS Total Mood Disturbance (TMD) delta from baseline scores by treatment condition and week of training (error bars are for 99% confidence intervals). 435
Figure VI-27. Estimated marginal means for POMS T-factor delta from baseline scores by chronotype for actigraphy subsample (error bars are for 99% confidence intervals). 437
Figure VI-28. Estimated marginal means for POMS A-factor delta from baseline scores by chronotype for actigraphy subsample (error bars are for 99% confidence intervals). 437
Figure VI-29. Estimated marginal means for POMS V-factor delta from baseline scores by chronotype for actigraphy subsample (error bars are for 99% confidence intervals). 438
Figure VI-30. Estimated marginal means for POMS C-factor delta from baseline scores by chronotype for actigraphy subsample (error bars are for 99% confidence intervals). 438
Figure VI-31. Estimated marginal means for POMS D-factor delta from baseline scores by treatment condition and chronotype for actigraphy subsample (error bars are for 99% confidence intervals). 439
Figure VI-32. Estimated marginal means for POMS V-factor delta from baseline scores by treatment condition and chronotype for actigraphy subsample (error bars are for 99% confidence intervals). 440
Figure VI-33. Estimated marginal means for POMS Total Mood Disturbance (TMD) delta from baseline scores by week of training for the actigraphy subsample (error bars are for 99% confidence intervals). 442
Figure VI-34. Estimated marginal means for POMS Total Mood Disturbance (TMD) delta from baseline scores by chronotype for actigraphy subsample (error bars are for 99% confidence intervals). 443
Figure VI-35. Estimated marginal means for POMS Total Mood Disturbance (TMD) delta from baseline scores by study condition and chronotype for actigraphy subsample (error bars are for 99% confidence intervals). 444
Figure VI-36. Estimated marginal means for first and last marksmanship scores by treatment condition (error bars are for 95% confidence intervals). 448
Figure VI-37. Estimated marginal means for first and last marksmanship scores by treatment condition for the actigraphy subsample (error bars are for 95% confidence intervals). 451
Figure VI-38. Estimated marginal means for push-up score by week of training (error bars are for 99% confidence intervals). 456
Figure VI-39. Estimated marginal means for push-up score by treatment condition and week of training (error bars are for 99% confidence intervals). 457
Figure VI-40. Estimated marginal means for sit-up score by week of training (error bars are for 99% confidence intervals). 459
Figure VI-41. Estimated marginal means for sit-up score by chronotype (error bars are for 99% confidence intervals). 460
Figure VI-42. Estimated marginal means for sit-up score by treatment condition and week of training (error bars are for 99% confidence intervals). 461

Figure VI-43. Estimated marginal means for run score by week of training (error bars are for 99% confidence intervals). 463
Figure VI-44. Estimated marginal means for run score by chronotype (error bars are for 99% confidence intervals)....................464
Figure VI-45. Estimated marginal means for Army Physical Fitness Test (APFT) score by week of training (error bars are for 99% confidence intervals). 467
Figure VI-46. Estimated marginal means for Army Physical Fitness Test (APFT) score by chronotype (error bars are for 99% confidence intervals). 468
Figure VI-47. Estimated marginal means for Army Physical Fitness Test (APFT) score by treatment condition and week of training (error bars are for 99% confidence intervals). 469
Figure VI-48. Estimated marginal means for ESS score by treatment condition and week of training (error bars are for 95% confidence intervals). 472
Figure VI-49. Estimated marginal means for ESS score by chronotype (error bars are for 95% confidence intervals)....................474
Figure VI-50. Estimated marginal means for PSQI score by treatment condition and pre/post-training (error bars are for 95% confidence intervals)....................476
Figure VI-51. Histogram of participants' ratings of their own sleep (top) and their peers' sleep (bottom) by treatment condition. 478
Figure VI-52. Histogram of daily frequency that participants report falling asleep during activities by treatment condition. 479
Figure VI-53. Histogram of participants' preferred time of the day for physical fitness training by treatment condition. 480
Figure VI-54. Isoperformance curves trading off aptitude, expressed as initial marksmanship score, and average daily sleep, setting the final marksmanship score criterion levels at 23, 27, and 30 and percentage proficient at 80%. 496
Figure VI-55. Isoperformance curves trading off baseline sleep quality, expressed as pre-training PSQI score, and average daily sleep, setting the final PSQI score criterion levels at 5 and 6.5 and the assurance level at 80%. 500
Figure VII-1. Block diagram of a generic dynamic system. 511
Figure VII-2. Stock and flow diagram of the SAFTE model. 512
Figure VII-3. Task Effectiveness Scheduling Tool results when the predicted task effectiveness criterion is set to 95% and sleep quality is rated as good. 520
Figure VII-4. Task Effectiveness Scheduling Tool results when the predicted task effectiveness criterion is set to 90% and sleep quality is rated as good. 521
Figure VII-5. Task Effectiveness Scheduling Tool results when the predicted task effectiveness criterion is set to 90%, sleep quality is rated as good, and a 10 hours-of-work rule is enforced....................522
Figure VII-6. Task Effectiveness Scheduling Tool results when the predicted task effectiveness criterion is relaxed to the FAST criterion line of 77.5% and sleep quality is rated as poor. 524
Figure IX-1. A family of interacting systems, to include a system-of-interest and its sibling systems, all existing within the environment provided by the containing system....................544

Figure IX-2. Interacting systems readjusting to the addition of a new system-of-interest [From Hitchins, 1992]....545
Figure IX-3. Generic reference model and New HSI [After Hitchins, 1992]....552

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

Table II-1. Comparison among three large-system approaches [After Kleiner, 2008]......23
Table II-2. Machine Age versus Systems Age paradigms [From Ackoff, 1981]. 25
Table II-3. Boulding’s classification of systems [After Boulding, 1956].........................31
Table II-4. Practical aspects of hard and soft systems approaches [From Pidd, 2004]. 42
Table II-5. The main attributes of five metaphors used in the creativity phase of TSI [From Flood, 1995]. 49
Table II-6. Characteristics of simple and complex systems [After Flood & Jackson, 1991]. 51
Table II-7. Characteristics of unitary, pluralist, and coercive relationships between participants [After Flood & Jackson, 1991]..52
Table IV-1. Variables in the power law of practice and corresponding HSI training and personnel domain considerations. 198
Table IV-2. Physical programming region limits table. 229
Table IV-3. Physical programming optimization results..229
Table IV-4. Raw data matrix for airline passenger screening systems. 232
Table IV-5. 2 x 3 study design matrix. 236
Table IV-6. Initial results for the airline passenger screening system experiment...........237
Table IV-7. Physical programming region limits table. 250
Table IV-8. Physical programming optimization results..251
Table V-1. Estimated regression coefficients and standard errors for the final fitted model of the basic maneuvering task data. 301
Table V-2. Estimated regression coefficients and standard errors for the final fitted model of the landing task data. 308
Table V-3. Estimated regression coefficients and standard errors for the final fitted model of the landing task data. 314
Table V-4. Estimated regression coefficients and standard errors for the final fitted model of the reconnaissance task data. 317
Table V-5. Personnel cost estimates. 327
Table V-6. Residuals for the basic maneuvering task data. 341
Table V-7. Residuals for the basic maneuvering task data (continued). 342
Table V-8. Residuals for the basic maneuvering task data (continued). 343
Table V-9. Residuals for the landing task data. 351
Table V-10. Residuals for the landing task data (continued). 352
Table V-11. Residuals for the reconnaissance task data. 366
Table V-12. Residuals for the reconnaissance task data (continued). 367
Table VI-1. Summary of study variables..377
Table VI-2. Schedule for data-generating events. 393
Table VI-3. Summary of intervention and comparison study groups at outset of study. ..396
Table VI-4. Summary of intervention and comparison study groups at outset of study (continued). 397
Table VI-5. Summary of intervention and comparison study groups at outset of study (continued). 398

Table VI-6. Summary of intervention and comparison study groups for actigraphy subsample at outset of study. 402
Table VI-7. Summary of intervention and comparison study groups for actigraphy subsample at outset of study (continued). 403
Table VI-8. Summary of intervention and comparison study groups for actigraphy subsample at outset of study (continued). 404
Table VI-9. Univariate tests for weekly average sleep. 409
Table VI-10. Population moment estimates based on sample k^{th} moments. 414
Table VI-11. Multivariate tests for POMS delta from baseline scores. 418
Table VI-12. Univariate tests of between-participant effects for POMS T-factor delta from baseline scores. 419
Table VI-13. Univariate tests of between-participant effects for POMS D-factor delta from baseline scores. 421
Table VI-14. Univariate tests of between-participant effects for POMS A-factor delta from baseline scores. 423
Table VI-15. Univariate tests of between-participant effects for POMS V-factor delta from baseline scores. 426
Table VI-16. Univariate tests of between-participant effects for POMS F-factor delta from baseline scores. 428
Table VI-17. Univariate tests of between-participant effects for POMS C-factor delta from baseline scores. 431
Table VI-18. Univariate tests for Total Mood Disturbance delta from baseline scores. 433
Table VI-19. Multivariate tests for POMS delta from baseline scores for actigraphy subsample. 436
Table VI-20. Univariate tests of between-participant effects for Total Mood Disturbance delta from baseline scores for actigraphy subsample. 441
Table VI-21. Within-participant effects for marksmanship score. 446
Table VI-22. Within-participant effects for marksmanship score (continued). 447
Table VI-23. Between-participant effects for marksmanship score. 449
Table VI-24. Between-participant effects for marksmanship score (continued). 450
Table VI-25. Within-participant effects for marksmanship score for the actigraphy subsample. 451
Table VI-26. Between-participant effects for marksmanship score for the actigraphy subsample. 452
Table VI-27. Multivariate tests for physical fitness component scores. 453
Table VI-28. Multivariate tests for physical fitness component scores (continued). 454
Table VI-29. Univariate tests of between-participant effects for push-up score. 455
Table VI-30. Univariate tests of between-participant effects for sit-up score. 458
Table VI-31. Univariate tests of between-participant effects for run score. 462
Table VI-32. Univariate tests of between-participant effects for Army Physical Fitness Test score. 465
Table VI-33. Univariate tests of between-participant effects for Army Physical Fitness Test score (continued). 466
Table VI-34. Within-participant effects for Epworth Sleepiness Scale score. 471
Table VI-35. Between-participant effects for Epworth Sleepiness Scale score. 473

Table VI-36. Within-participant effects for Pittsburgh Sleep Quality Index score............475
Table VI-37. Between-participant effects for Pittsburgh Sleep Quality Index score.........477
Table VI-38. Results for the fitted binary logistic regression model for failure to graduate with initial training cohort...482
Table VI-39. Expected mean squares, F ratio, significance level, and effect size for the basic rifle marksmanship data from the actigraphy subsample. 493
Table VI-40. Expected mean squares, F ratio, significance level, and effect size for the post-training PSQI score data from the actigraphy subsample......................498
Table IX-1. New Human Systems Integration guidelines...542

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ABBREVIATIONS AND ACRONYMS

CLD	causal loop diagram
DOE	design of experiments
DP	design points (DP)
ESS	Epworth Sleepiness Scale
FAST	Fatigue Avoidance Scheduling Tool
HFE	human factors engineering
HSI	human systems integration
INCOSE	International Council of Systems Engineering
LP	linear programming
NDI	non-developmental item
NIB	nominal-is-better
OVO	One versus Others criteria rule
PSQI	Pittsburgh Sleep Quality Index
REM	rapid eye movement
RFP	request for proposals
SAFTE	Sleep, Activity, Fatigue, and Task Effectiveness Model
SIB	smaller-is-better
SOI	system-of-interest
SWS	slow wave sleep
TEST	Task Effectiveness Scheduling Tool
TSA	Transportation Security Administration

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The major purpose in undertaking this discourse was to tackle two fundamental questions: *What is human systems integration (HSI) and how should one think about HSI problems?* The objective was to develop a coherent systems method that would improve the integration of the HSI domains to create sustainable systems while preserving consideration of system stakeholder preferences. Addressing these questions first required putting the concept of HSI in context, both in terms of a philosophy and a Defense Department program.

The lesson learned from the juxtaposition of these two conceptual views (i.e., philosophy versus program) was the *rejecting of the notion that HSI is simply "post-modern" human factors*. HSI as a philosophy evolved within the context of the larger systems movement that occurred in the 1960s in response to the issue of irreducible complexity. HSI emerged in response to real-world, macroergonomic, political and military challenges that resulted in an organizational crisis. This crisis, in the simplest of terms, was caused by technological complexity and its effects on personnel. Thus, the fundamental impetus for HSI was *complexity*.

Allowing philosophy to inform method, the lessons learned from the historical analysis were used to characterize and illustrate an approach to addressing HSI issues early in a weapon system acquisition process. The following prime directive—the highest level of abstract, objective statement of purpose—was proposed for an HSI program: *To produce sustained system performance that is humanly, technologically, and economically feasible.* Based on an analysis of this prime directive, and with an implicit reference to sociotechnical systems theory, the following definition of HSI was derived:

> A philosophy applied to personnel and technological subsystems within organizations in pursuit of their joint optimization in terms of maximally satisfying organizational objectives at minimum life cycle cost. Its practice is concerned with the specification and design for reliability, availability, and maintainability of both the personnel and technological subsystems over their envisioned life cycle.

We assert that the principle approach to HSI should involve the integration of the behavioral sciences, human factors engineering, and operations research to more broadly represent human considerations early in weapon system analyses and in the products that evolve from these analyses.

Inclusion of HSI in system analyses necessitates a holistic perspective of the trade space formed by the synthesis of economic considerations and the individual HSI domains and their interactions. This conceptualization of HSI was expanded to include both a macro-HSI and micro-HSI trade space. The goal of HSI then becomes one of ensuring that micro HSI tradeoffs are organizationally net positive. Ideally, the micro-HSI trade space can be considered in systems analyses by integrating Simon's research strategy of efficient multifactor design of experiments, Kennedy and Jones' isoperformance approach, and coupling isoperformance with utility analysis through means such as physical programming.

Three case studies were used to illustrate this paradigm of integrating the behavioral sciences and human factors engineering with the tools and methodologies of operations research to address HSI issues. The first case study used an opportunistic dataset from a USAF study evaluating the impact of prior flight experience on acquisition of unmanned aircraft system operator skills. Isoreliability models were then constructed and aggregated across system functions, thereby allowing consideration of personnel and training domain tradeoffs in terms of total system reliability. The second case study applied the isoperformance methodology to data from a prospective study examining the effect of a sleep scheduling intervention on measures of Soldier performance during Basic Combat Training. Tradeoff models were constructed for both rifle marksmanship performance and occupational health in terms of the personnel and survivability domains of HSI. The third case study used a mixed integer program—the Task Effectiveness Scheduling Tool—to analyze simulation data derived from a validated biomathematical fatigue model to explore the trade space that exists between the manpower, survivability, habitability, and human factors engineering domains of HSI.

Finally, based on a meta-synthesis of the aforementioned concepts and ideas, design guidelines for a New HSI method were proposed and future challenges discussed.

ACKNOWLEDGMENTS

I would never have been able to finish my dissertation without the guidance of my committee members and the support from my family.

I would like to thank my advisor, Dr. Nita Shattuck, for her guidance and for providing me with an excellent environment for doing research.

Most importantly, I would like to express my deepest gratitude to my beloved wife, Michelle, and my son, Jonathon, who both waited patiently and supported my work.

Finally, I wish to specifically call out William Blackwood, Robert Kennedy, Marshall Jones, and Charles Simon, the last unfortunately being deceased. The several hours of interviews I conducted with Lieutenant Colonel Blackwood (USA, retired), along with the follow-up e-mails, were absolutely critical to my piecing together the history of the Army's MANPRINT program, and I greatly appreciated the opportunity to learn this history from his first person perspective. I also owe a sincere debt of gratitude to Kennedy, Jones, and Simon, on whose collective shoulders I was able to figuratively stand in my effort to develop this dissertation. I must thank Dr. Kennedy for sending me his personal collection of papers on isoperformance, which also contained many of Simon's technical reports. Similarly, I wish to thank Dr. Jones for his review and critique of major parts of this dissertation manuscript.

There are many other people who made significant contributions to my dissertation, either directly in terms of the conduct of the research itself or indirectly through my postgraduate education. While it is not possible to enumerate the names of all of these people here, I nevertheless want to offer each and every one my sincerest thanks.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

> To a very great degree, all of us are products of our experiences. We are products of our own times and our own experiences. We accept as "truth" only those wisdoms that our own experience validates as being true. I would encourage you…to recognize that you will not have had an opportunity to experience all of those things that your colleagues have. You will not be able to validate, by your own experience, all of the truths that maybe they have validated by theirs (General Russell E. Dougherty, USAF, retired, 1992)

A. ORIENTATION

If, like me, you have spent much time working in the human performance-related domains, systems engineering, or defense systems acquisition, you may have noticed that there is an incomplete understanding of Human Systems Integration (HSI). Such ambiguity can be attributed, in no small part, to the lack of a general consensus on the definition, scope, and intent of HSI as well as the corresponding body of knowledge it is supposed to cover. HSI practitioners have a problem, not only explaining to other people what they do, but also defining it amongst themselves. This problem is further exacerbated by the internal fragmentation of the HSI work force according to vocational specialty and educational background. Reflecting on personal experience and reviewing the literature, there does not appear to be a unique body of knowledge for HSI. The individual HSI domains are disciplines and careers in themselves and each has its own literature. Consequently, individuals charged with integrating the HSI domains often lack both a clear mental image of the trade space and a set of basic principles that they can put to useful work. This is evidenced by my frequent observation that graduate students in HSI at the Naval Postgraduate School, when asked, struggle to illustrate a simple conceptual model of HSI—a difficulty shared by many HSI practitioners, program managers, and engineers. Thus, we are left to collectively ponder the nagging question, "what is HSI, and how should it work?"

Our incomplete understanding of HSI is a problem because lots of dedicated people are spending energy and resources trying to develop educational programs and courseware, promulgate policies, and create tools to conduct HSI. Worse yet, program

managers, system engineers, and HSI practitioners are presently grappling with integrating "the human element" into "technological systems" and becoming greatly confused in the process. They have difficulty incorporating the spectrum of HSI considerations in their decision making since they lack a systematic process for "pulling together" domain-specific studies and expertise in an applied situation. Yet integration of the HSI domains will inevitably occur in virtually all system acquisitions. That is, domain interactions will occur in an *ad hoc* fashion rather than by deliberate design. The consequences of this approach may include prohibitive total ownership costs and failure to attain system performance thresholds. Even when system performance objectives are met, there are lost opportunity costs resulting from decreased productivity and wasted resources.

While admirable work is being accomplished by various HSI stakeholders, much of it is focused on progressively breaking down HSI into ever more detailed sets of technical and engineering management activities. In so doing, HSI proponents have presented their approach as logical, rational, and multi-disciplinary, but in the whole, it is seemingly not based on any science in the way that the constituent HSI domains were, such as the human factors engineering or training domains. Echoing the charge proffered by Hitchins (1992) against systems engineers, I assert that instead of HSI theory there has developed a HSI "theology." Part of the HSI theology includes the development of design options, assessment of the individual HSI domain considerations for these design options and their subsequent tradeoff to select the optimal solution in terms of total system performance and ownership cost. Trading between HSI domains is, at best, a crude art as presently practiced and there is no agreed upon metric for defining what "optimal" means in terms of HSI. All in all, it can be said that the theology on which HSI is supposed to be based has dubious foundations in policy guidance that has evolved over the past two decades. Nevertheless, suggestions to create a more robust philosophical or scientific underpinning for HSI continue to receive little approbation from either HSI proponents or detractors.

For the skeptical reader, I will borrow an example from Hamming (1997) to illustrate the challenges of such "faith-based" thinking about the HSI trade space.

Putting off any discussion of HSI definitions for the moment, the Defense Department describes HSI in terms of seven factors or domains (Department of Defense, 2008), which by implication means that we should describe a HSI solution in terms of at least seven parameters and other military services expand this number to eight, nine, or even higher. Hence, while we may build and operate systems in 3-dimensional space, system designers and HSI practitioners must be concerned with a higher (*N*) dimensional design space that has one dimension for each design parameter.

Although *N*-dimensional space is a mathematical construct, we must think about it to better understand what happens to us when we wander there with a HSI problem. To do so, we begin with a simple geometric example and consider a square whose edges are each four units in length (Figure I-1) and in which we place four unit circles (depicted in black), each circle having a radius of one unit. We then draw a circle (depicted in red) about the center of the square with radius just touching the four unit circles. Based on simple application of Pythagorean's theorem, its radius must be

$$r_2 = \sqrt{2} - 1 = 0.414\ldots$$

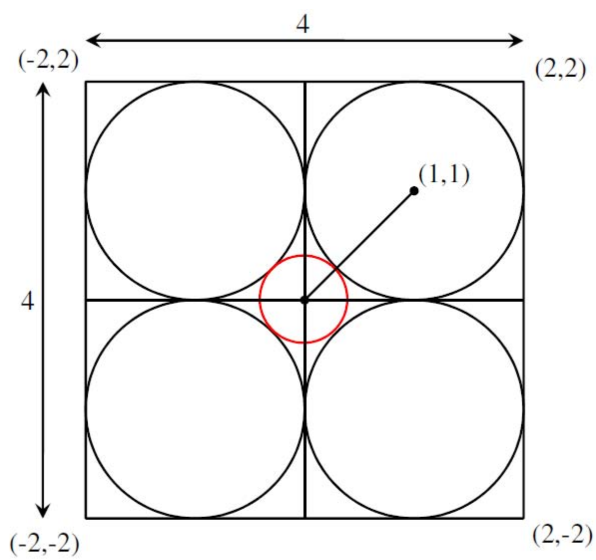


Figure I-1. Balls enclosing ball [After Hamming, 1997].

In three dimensions, we have a cube whose edges are each four units and into which we can place eight spheres of unit radius. Now consider a sphere circumscribed

about the center of the cube with radius just touching the surface of the other spheres. By repetitive application of Pythagorean's theorem, this sphere must have a radius of

$$r_3 = \sqrt{3} - 1 = 0.732\ldots$$

By induction, as we go to n dimensions, we will have a $4 \times 4 \times \ldots \times 4$ cube with 2^n spheres, one in each corner of the cube and touching its n adjacent neighbors. A sphere circumscribed at the center of this cube with radius touching the surface of the other spheres will have a radius of

$$r_n = \sqrt{n} - 1$$

Now consider the case of ten dimensions where the radius of the central sphere is

$$r_{10} = \sqrt{10} - 1 = 2.162\ldots > 2$$

In ten dimensions, the radius of the central sphere reaches outside the surrounding cube—an apparent paradox!

The point of this example is that simple raw intuition is inadequate when considering *N*-dimensional space. However, this is the very space where the design of HSI solutions generally takes place. As stated by Hamming, it is not 3-dimensional space that matters in system design, but rather it is *N*-dimensional space, and *N*-dimensional space can be very vast. To illustrate the latter, consider for a moment an HSI problem in which the set of potential solutions is limited to only two design options per domain and a proper solution requires a choice be made for every domain. In this relatively constrained scenario, there are 2^7 or 128 solutions to consider if we entertain seven domains and 2^9 or 512 solutions with nine domains. If we relax our constraints and allow ten potential design choices per domain, we have ten million solutions to consider. Even if you could model and compute all these design solutions, there would be insufficient time to even look through them, let alone test them! Clearly, we cannot mechanically explore the HSI trade space using generic tools without some form of educated inspiration. This then leads me to the following proposition:

> Accommodating the human element in technological systems is an *N*-dimensional creativity problem, not a 3-dimensional ergonomics problem.

By considering HSI from the perspective of creative design or problem solving in *N*-dimensional space, it becomes clear that accommodating the human element is a very hard problem. If one makes a small change in one HSI variable, it tends to reverberate throughout the entire system, often times with unintended consequences. This makes even small perturbations potentially difficult to cope with. I will illustrate the concept in a HSI context using a causal loop diagram (CLD) or influence diagram derived from the work by Miller and Firehammer (2007). For the novice, a CLD is a systems thinking tool that depicts a diagram with arrows connecting variables (i.e., things that change over time) in a way that shows how one variable affects another.

Now consider the simple CLD depicted in Figure I-2 showing the potential implications of changes in the numbers of human resources (i.e., manpower) provided to operate and maintain a military system. Manpower related costs significantly drive a system's total life-cycle cost and can be as much as 80% of total operations and support costs (U.S. Air Force, 2008). It should thus come as no surprise that senior decision makers and system designers often look for opportunities to reduce manpower when developing new systems or upgrading legacy systems. However, requirements to reduce manpower frequently result in system designers allocating more tasks and roles to individual crewmembers with consequent increases in their overall workload. Increased workload, when not mitigated by adequate opportunities for rest and recovery, results in chronic fatigue, which subsequently leads to decreased productivity and increased risk of errors and mishaps. These outcomes, in turn, drive up life-cycle costs in contrast to the system designer's original expectations. Alternatively, the system's owners may later opt to provide opportunities for recovery through schedule changes, but such changes require increased manpower, and consequently, increased life-cycle costs.

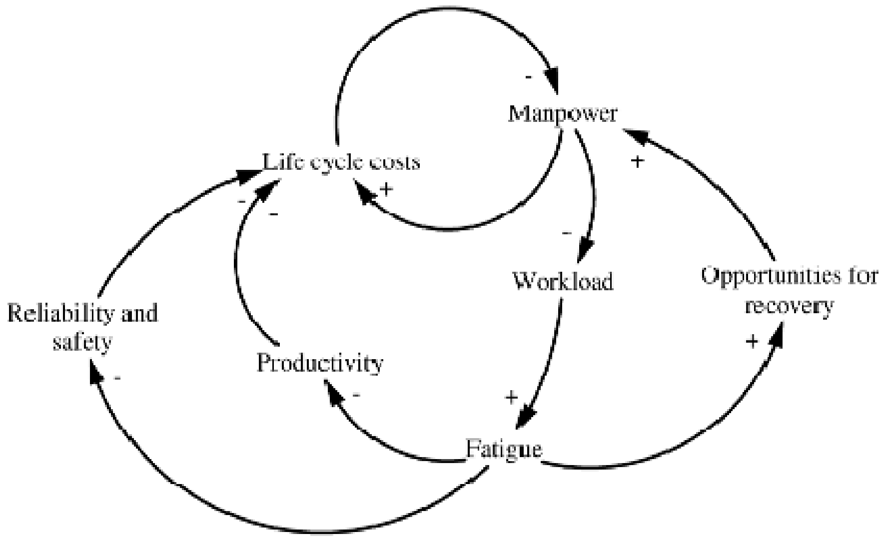


Figure I-2. Causal loop diagram showing the implications of manpower changes.

Hopefully the casual reader has not been put off by the discussion of such an abstract construct as *N*-dimensional space; it was mainly a prop used to illustrate the innate complexity of the HSI trade space that all too often seems to go unappreciated, perhaps even unrealized, by many self-proclaimed HSI experts. Nevertheless, as a theology, HSI continues to survive principally because it provides a way of approaching the human element in systems that appears axiomatically sound. Absent demonstrable evidence of recurring success, the current approach to HSI invites reference to the popular refrain on insanity as "doing the same activity over and over and expecting different results." Hence, in developing the ideas for this dissertation, I perceived the need for a more systemic and systematic approach to HSI than is presently discernable within the Defense Department's integrated lifecycle management framework. Such an approach should promote a more harmonious balance in considering human capabilities versus technological capabilities early in system acquisition, while simultaneously preserving the focus on system stakeholder values in identifying preferred solutions. The topics and objective of this dissertation are shown in Figure I-3, which has been organized into an intent structure starting with foundation and theory at the bottom and

culminating in a future vision at the top. Such a structure is often used to develop mission statements for organizations, and so my mission statement is:

> To develop a coherent systems method that will improve the integration of the HSI domains to create sustainable systems while preserving consideration of system stakeholder preferences.

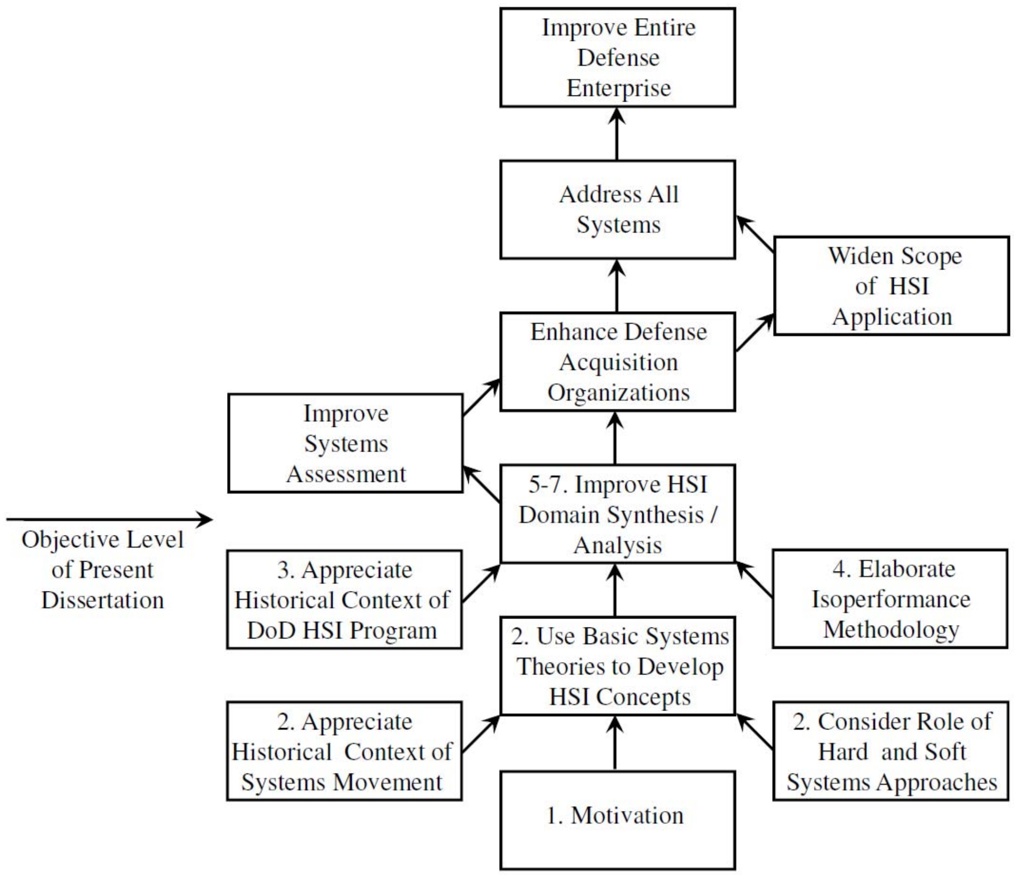


Figure I-3. Intent structure outlining topics and objectives starting with foundation and theory at the bottom and working up to a future vision at the top. Numbered boxes correspond to dissertation chapters.

Figure I-4, intended as a guide or road map, shows my approach to form a bridge from a HSI process-oriented theology to a more enlightened state of understanding. The first step is to identify the vague, unstructured issues underlying the need for HSI. My objective is to move from these vague issues progressively towards solutions, borrowing

along the way from substantiated work in related fields where applicable. In particular, although perhaps not qualifying as a substantiated work, Derek Hitchins' *Putting Systems to Work* (1992) plays a significant role in framing my thinking and approach to HSI and those familiar with his book will note some parallel themes. My desired end state, and hence benchmark for success, is an intellectually coherent and defensible architecture for relating HSI domain considerations that addresses the issues for which HSI was originally devised.

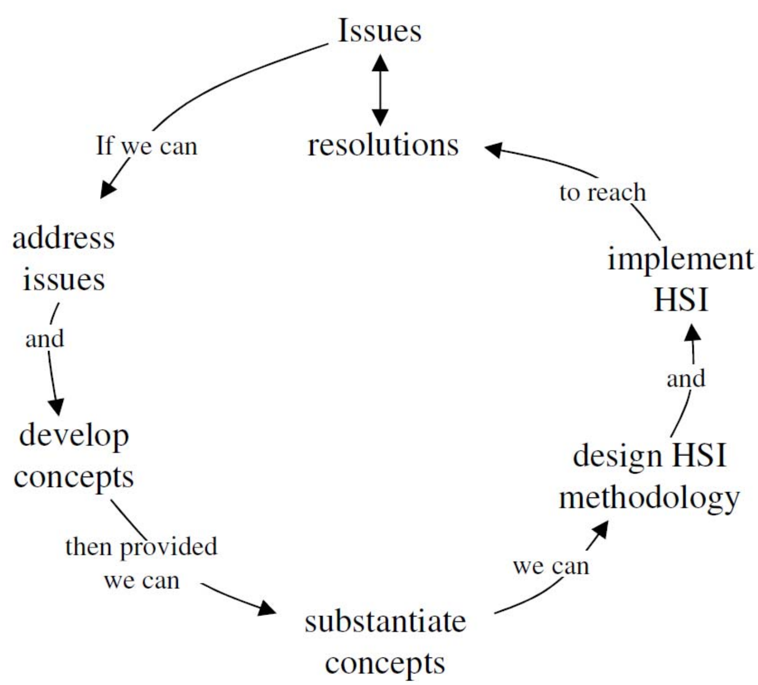


Figure I-4. Bridging from issues to HSI solutions [After Hitchins, 1992].

My intent in writing this dissertation is for it to be useful for systems practitioners responsible for addressing HSI problems and issues. I am as much concerned about displaying a way of thinking about HSI problems as in advancing any particular technical approach. Since it is questionable whether a way of thinking can be conveyed simply by narrative description, it is my intent to approach the topic through examples in the form of the studies described in Chapters V, VI, and VII. You should need no special skills to

understand this work other than perhaps an introductory experience in operations research or systems engineering. Although at times I resort to mathematics to illustrate and connect ideas and concepts, I have strived to ensure that the underlying ideas can be grasped from the words alone. Much of this work has been developed from both my experiences as a HSI practitioner and the constellation of insights garnered during lectures and projects in my postgraduate studies. I have deviated from the normal dissertation format to provide a concise presentation of topics that should be useful for experienced systems practitioners with no formal qualifications in HSI.

B. BIBLIOGRAPHY

Department of Defense (2008). *Department of Defense instruction 5000.02: operation of the defense acquisition system*. Washington, DC: Author.

Dougherty, R. E. (1992, Summer). Roots and wings: a perspective on reorganization. *Airpower Journal*, VI(2). Retrieved June 25, 2010, from http://www.airpower. maxwell.af.mil/airchronicles/apj/apj92/sum92/doughty.htm

Hamming, R. W. (1997). *The art of doing science and engineering: learning to learn*. Amsterdam B.V., The Netherlands: Gordon and Breach Science Publishers.

Hitchins, D. K. (1992). *Putting systems to work.* Chichester, England: John Wiley and Sons.

Miller, N. L., & Firehammer, R. (2007). Avoiding a second hollow force: the case for including crew endurance factors in the afloat staffing policies of the U.S. Navy. *Naval Engineers Journal*, 1, 83–96.

United States Air Force (2008). *Human systems integration development guide: capability based requirements.* Washington, DC: Government Printing Office.

THIS PAGE INTENTIONALLY LEFT BLANK

II. HUMAN SYSTEMS INTEGRATION PHILOSOPHY

> The professionalized cognitive and occupational styles that were refined in the first half of this century, based in Newtonian mechanistic physics, are not readily adapted to contemporary conceptions of interacting open systems and to contemporary concerns with equity (Rittel & Webber, 1973, p. 13)

A. AN INTRODUCTION TO HUMAN SYSTEMS INTEGRATION

As of this writing, there is no general consensus on the definition of HSI as evidenced by a white paper prepared by the International Council of Systems Engineering (INCOSE) HSI Working Group that identified 39 definitions (Deal, 2007). So where to start? While it is not our purpose here to join this debate, let us delve into a few general ideas. For instance, if we are to think about HSI, what do the constituent words mean? A "human" is simply a bipedal primate mammal, or so the dictionary definition states (Merriam-Webster, 2009). The definition of a "system" is somewhat more abstract (Merriam-Webster, 2009):

- A regularly interacting or interdependent group of items forming a unified whole
- An organized set of doctrines, ideas, or principles usually intended to explain the arrangement or working of a systematic whole
- An organized or established procedure; a manner of classifying, symbolizing, or schematizing
- Harmonious arrangement or pattern.

And lastly, the definition of "integrate" (i.e., the verb form of integration) includes (Merriam-Webster, 2009):

- To form, coordinate, or blend into a functioning or unified whole
- To find the integral of (as a function or equation)
- To unite with something else; to incorporate into a larger unit
- To end the segregation of and bring into equal membership in society or an organization.

Notwithstanding that the definition of "system" seems all-embracing, the combinatorial sum of these definitions can give rise to a variety of divergent viewpoints

of HSI. For example, one could describe HSI as incorporating bipedal primate mammals into an interacting group of items to form a unified whole. If those items are technological artifacts, say workstations, one develops an engineering-centric viewpoint of HSI. If those items are instead doctrines or policies, then the perspective changes to that of organizational behavior and the management sciences. There is also a moralistic viewpoint if one chooses to define integration in terms of equal membership. Believing there is an organizational or societal tendency to overemphasize the technological elements of systems, one might argue for coequal consideration of humans in systems. Lest you think this moralistic interpretation is overreaching:

> …our "equipment" oriented culture needs to change to one that is "people" oriented (Booher, 1990, p. 2).
>
> As science, [human factors engineering] is needed to understand the ramifications of the human-technology relationship. Note that the word *human* [emphasis in original] precedes *technology*; that is because technology should be the servant, not the master, although in too many instances the roles are reversed (Meister, 1999, p. 359).

One might even envision such a moralistic imperative evolving into a formal legal viewpoint of HSI should self-aware, intelligent machines ever be realized (Brooke, 2009). Finally, there is precedent for the mathematical definition of integration, as in finding the integral, being used to describe HSI, at least symbolically, in Booher's (2003) double-integration process model (Figure II-1).

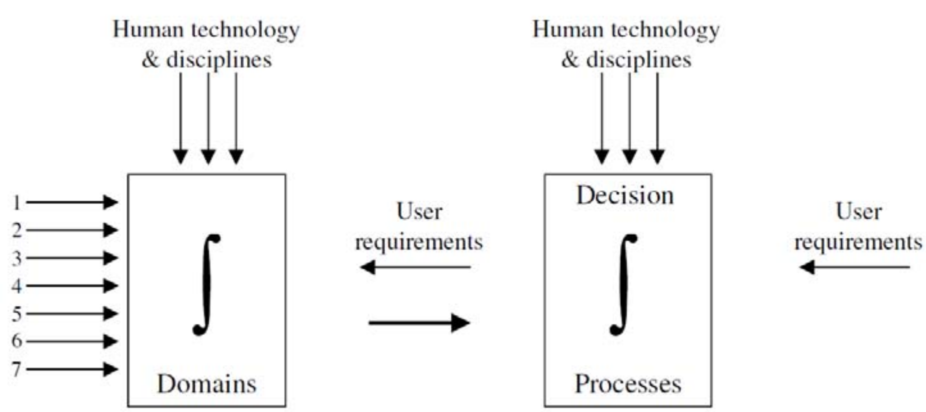


Figure II-1. HSI double-integration process model [From Booher, 2003].

Without appealing to any of the potentially competing camps of HSI stakeholders, we have just generated several definitions of HSI that give rise to a variety of viewpoints, or *Weltanschauungen*—"worldviews" in English. These include:

- An engineering view
- An organizational behavior and management view
- A moral view
- A mathematical view.

It is reassuring that Deal (2007) independently makes a similar observation in his survey of definitions for the INCOSE HSI Working Group:

> [HSI] definitions refer to the same target concept, but from different perspectives. Therefore, HSI definitions show a significant scatter of concept and element (p. 2).

The idea of developing different *Weltanschauungen* when exploring a problem situation is a keystone of modern systems thinking (Checkland, 2000). If definitions are being driven by viewpoints, then it is counterproductive to try to identify a single, comprehensive definition for HSI. Unfortunately, the stated intent of the INCOSE HSI Working Group is to formulate just such a universal definition (Deal 2007, p. 1). The result of their work, not surprisingly, is a definition that captures the engineering viewpoint:

> Human systems integration is the interdisciplinary technical and management processes for integrating human considerations within and across all system elements; an essential enabler to systems engineering management (Mueller, 2008, p. 7).

To the contrary, we should entertain the idea that a portfolio of HSI definitions, borne from a diversity of viewpoints, is probably desirable. For example, von Bertalanffy (1972), an early proponent of general systems theory, views multiple definitions as a healthy development in a new field:

> The existence of different descriptions is nothing extraordinary and is often encountered in mathematics and science… (p. 415).

Deal (2007) notes that the Defense Department's policy guidance for operation of its acquisition system is the source document for many HSI definitions. This policy

guidance is a physical artifact of the Defense Department management's efforts to address a particular problem situation that evolved over several decades (Booher, 2003). It defines the objectives and scope and delineates accountability for the HSI program (Department of Defense [DoD], 2008) – it describes the Defense Department's way of doing business with respect to HSI. What it does not do is address how to think about HSI as a systems discipline. For that, it is necessary to distinguish "program" from "philosophy," which brings us again to the moral viewpoint of HSI.

So, from whence did "HSI philosophy" come? Clearly, we cannot formulate a strictly objective answer to this question, for as Popper (1957) asserts, the best we can hope to accomplish is to write a history that is consistent with a particular point of view. Thus, we should, if possible, clearly articulate that point of view. Accordingly, our intent is to provide a sketch of the development of HSI that enables us to understand the nature of HSI as being complementary to the human factors sciences that formalized in the first half of the twentieth century. Checkland (1981) provides an excellent overview of the long story of the development of the science movement in Western civilization and the more recent emergence of the related systems movement, and we borrow heavily from his work to develop our ideas here. What follows, then, is only a brief outline of some of the more pertinent features of relatively recent efforts to address human performance in systems using the method of science. Its purpose is to provide an accounting of the initial attempts to address the complexity of the problem through reductionist thinking so that we can explain the emergence of HSI within the broader sweep of the systems movement.

Following from the seminal lecture series, *Lectures on Men and Machines*, by Chapanis, Garner, Morgan, and Sanford (1947), the foundations for a formal scientific discipline focused on humans in systems were laid during the Second World War:

> Only within the last few years have most of us realized how little of practical value is known about the coordination between man's senses and his muscles. During the war, this ignorance frequently cost us much in both men and materials. To fill some of the specific gaps in our knowledge, a great deal of psycho-physical research was started. One of the resulting efforts was the Harvard "Systems Research" contract…a project to improve the complex information systems which made up our

> shipboard Combat Information Centers…At the end of hostilities, "Systems Research" was continued through a contract between the John Hopkins University and the Special Devices Center, Office of Naval Research. It was as a part of this contract that the ten excellent lectures…were delivered to the Naval Postgraduate School, Annapolis, in the spring of 1947…These lectures, the *first coherent public discussion of their subject* [emphasis added], show the work which has been done and, even more vividly, the work which still needs doing (pp. v–vi).

The war was a watershed for technological America and the West in the twentieth century (Meister, 1999). Wartime developments in radar, communications, sonar, aircraft, and combat information centers significantly increased the complexity of the tasks required of the human and served to highlight the rising problem of human performance in systems (Chapanis et al., 1947):

> When we stop to think how much a single radar can do in a fraction of a second, and then stop to think also that even the simplest form of a reaction for a human being requires approximately 1.5 of a second, we realize the limitations we are up against. This simple comparison of a machine's reaction time with a man's reaction time furnishes us with a clear cut example of what we are up against. The human factor in any system must be studied. Machines that demand super-human performance will fail because the human is not yet in a super stage. Jobs that push man beyond his limits of skill, speed, sensitivity and endurance will not be done—cannot be done (pp. 12–13).

As described by Chapanis and colleagues (1947), appreciation of the problem of human performance in systems, in turn, led to the emergence of a new scientific discipline concerned with application of human factors to engineering design:

> The war needed, and produced, many complex machines, and it taxed the resources of both the designer and the operator in making them practical for human use. The war also brought together psychologists, physiologists, physicists, design engineers and motion-and-time engineers to solve some of these problems…Today there are many groups busy with research on man-machine problems. They use different names to describe the work in its various aspects…But whatever the name, the objective is the same—to develop, through fundamental research and applied tests, a science which can deal adequately with the design and operation of machines for human use (p. vii).

These early pioneers described their emerging field as "psychophysical systems research," later to be known as "human engineering," in a somewhat clumsy attempt to provide specificity regarding the types of systems that were the focus of their research, namely "systems of people and things" (Chapanis et al., 1947, p. 4). They also acknowledged a common lineage in time-and-motion engineering and experimental psychology, and while they offered that "personnel" and "educational" psychology were related to psychophysical systems research, they felt that these fields had developed into relatively distinct and independent branches of psychology. Such sentiment was made abundantly clear by Chapanis and colleagues (1947) in their very first lecture:

> There is no denying the very great significance of selection. In the general trend toward studying and doing something about the human being in his working environment, the studying of, and the doing something about, personnel selection has been, and is, of outstanding importance. We in the Systems Research Laboratory, however, are not primarily interested in this aspect of the total problem. We are interested in the man and the equipment with which he must work: we are interested in the design of the job, and we are interested in the design of the machine…We have chosen not to engage in systematic worry about the fact that one man may be better than another on a given job. There are many other people who are very competently worrying in this area (p. 10).

Hence, it appears that the problem of human performance in systems was parsed into distinct fields of inquiry almost from the very point in time at which there was cognizance of the problem. As described by Kennedy, Jones, and Balzley (1989) in their outline sketch of the progress of the broader human factors movement in the Defense Department, this arrangement established the pattern for the next four decades:

> Since the 1950's [sic] applied behavioral scientists working in the fields of systems, training, and selection have remained largely independent from each other. Within the Department of Defense, mission and function statements reinforce this separation. <u>Personnel</u> [emphasis in original] activities emphasize the use of correlational analyses. <u>Education</u> and <u>training</u> commands keep track of time course changes and employ repeated measures. In <u>Systems Research</u> and engineering psychology, man, machine and environmental interactions are studied. In systems work, the emphasis is often placed on the application of military standards and specifications. Methods include estimates (within probabilistic error

> boundaries) of central tendency and dispersion of human lawful relationships (transfer functions) from independent variable manipulations (p. 4).

Even within the field of psychophysical systems research, now called human factors engineering (HFE), specialization is a continuing trend:

> It may be that the trend to specialization is the most obvious change that one can discern. Whether there was ever a generic HFE may be a misconception because, in the early days, we were mostly working in a few specialty areas…Now that HFE has expanded into a much larger number of specialties (the latest [Human Factors and Ergonomics Society] Directory lists 20 specialty groups), it may appear that there is no common body of theory and methodology—or, more likely, that the commonality resides in the psychological background knowledge that most HFE professionals bring to the discipline (Meister, 1999, p. 206).

While it is commonplace today to bemoan such compartmentalization of knowledge, using terms like "stove-piped" or "siloed," we should by no means seek to discredit the thinking of the early pioneers in human factors. A cursory inspection of our world suggests that it may be fairly characterized as complex, being comprised of many parts that are densely connected. This was no less the case in the 1940s when early behavioral scientists were faced with the urgent wartime problem of addressing human performance in systems. They attempted to tackle this problem of complexity using the method of science, and in so doing, engaged the potent combination of rational thinking and experimentation that had demonstrably worked in the past.

Science copes with the complexity of the world by deconstructing phenomenon of interest into separate parts for analysis and study. Likewise, Checkland (1981) suggests that the practitioners of science manage complexity by dividing their knowledge of the world into different subjects or disciplines that are necessarily man-made and arbitrary. If we accept that our knowledge has to be arranged in this way because of our limited ability to take in the whole, then it is useful to arrange the classification of knowledge according to some rational principle. Many possible classifications may be proposed based on any number of different principles, and it is foolish to expect that any one version will be generally accepted given the different purposes for which a classification may be carried out. Perhaps the more important issue, as pointed out by

Checkland, is the apparent natural human tendency for these divisions to become so ingrained in our thinking that we begin to have difficulty seeing the unity that underlies the divisions.

As we reflect on the emergence of a human-systems discipline, it is useful to recall the classification of the sciences proposed by Auguste Comte (1865) in the 19th century as summarized by Checkland (1981):

> Comte's doctrine was that human thought in any subject area passed through three phases: a *theological* [emphasis in original] phase dominated by fetish beliefs and totemic religions; a *metaphysical* phase in which supernatural causes are replaced by 'forces', 'qualities', and 'properties'; and finally a *positive* phase in which the concern is to discover the universal laws governing phenomena…(p. 61).

Comte claimed that all sciences pass through this sequence. For example, chemistry progressed from alchemy to a positive science in the 18th century, and biology evolved during the 19th century from teleology (i.e., a doctrine that objects in the world fulfill their intrinsic nature or purpose) and vitalism (i.e., a doctrine that the processes of life are not explicable by the laws of physics and chemistry alone and that life is in some part self-determining) to a positive investigation of the laws relating living organisms in an environment. Similarly, human engineering (i.e., early human factors) turned its back on Darwinian trial and error to test the fit of the human to the machine (Meister, 1999) and began positive scientific investigations to determine "estimates of human lawful relationships from independent variable manipulations" (Kennedy, Jones, & Baltzley, 1988, p. 1) in the early 20th century. Comte's doctrine led him to place the sciences in a natural order, which with some updating by Checkland, assumes the following sequence: physics, chemistry, biology, psychology, and the social sciences, where physics is the most basic science, being concerned with mass, motion, force, and energy. Checkland notes that "the principles behind the classification are: the historical order of the emergence of the sciences; the fact that each rests upon the one which precedes it and prepares the way for the one which follows; the increasing degree of complexity of subject matter; and the increasing ease with which the facts studied by a

particular science may change” (pp. 61–62). The overarching pattern then is one of a hierarchy of levels of complexity that we find convenient to tackle through a hierarchy of separate sciences.

Checkland (1981) asserts that physics is most successful as a science because it best exemplifies the characteristics of the scientific method. These characteristics, which can be traced back to the history of the development of science, are reductionism, repeatability, and refutation: “We may *reduce* [emphasis in original] the complexity of the variety of the real world in experiments whose results are validated by their *repeatability*, and we may build knowledge by the *refutation* of hypotheses” (p. 51). Checkland then proceeds to raise the question of how the scientific method copes with increasingly complex problems beyond those encountered in physics. The main puzzle for him is that a new problem is seen to be a problem of that science and of the particular level of phenomena with which that science deals. For example, a phenomenon in chemistry can be explained in terms of the physics of the constituent atoms and molecules, such as their masses, energies, and force fields. However, this explanation does not explain away the fact that the phenomenon of chemistry exists or that it is capable of being investigated experimentally at a higher level of complexity than that encountered in physics. While physics can provide an accounting of the mechanism of some chemical phenomenon, it cannot explain the existence of problems of chemistry as such. If the latter were possible, we could reduce the science of chemistry to physics and simply address the problems at this lower level of complexity. Similarly, the problems of genetics and heredity, although explainable in terms of the chemistry of nucleic acids, are nevertheless problems of biology, and the science of biology cannot be explained away by, or collapsed into, the science of chemistry. Checkland thus concludes that each level of scientific complexity is characterized by its own autonomous problems. Moreover, he suggests that “the existence of the problem of the emergence of new phenomena at higher levels of complexity is itself a major problem for the method of science, and one which reductionist thinking has not been able to solve” (p. 65).

This concept of irreducible complexity may be easier to understand in terms of the “turtle metaphor” popularized by Stephen Hawking (1988):

> A well-known scientist (some say it was Bertrand Russell) once gave a public lecture on astronomy. He described how the earth orbits around the sun and how the sun, in turn, orbits around the center of a vast collection of stars called a galaxy. At the end of the lecture, a little old lady at the back of room got up and said: "What you have told is rubbish. The world is really a flat plate supported on the back of a giant tortoise." The scientist gave a superior smile before replying, "What is the tortoise standing on?" "You're very clever, young man, very clever," said the old lady. "But it's turtles all the way down!" (p. 1).

The metaphor illustrates the problem of the infinite regression argument, where each explanation requires a further explanation. Such an argument eventually leads to the problem of first causality, where a circular cause and consequence cycle occurs, this being an infinite tower of turtles in Hawking's example. In the case of epistemology, absent irreducible complexity, we would experience a similar regression argument, where a social phenomenon can be explained in terms of a phenomenon of psychology, and in turn, as a phenomenon of biology, chemistry, physics, and so on. Irreducible complexity explains why this is the not the case, because a phenomenon observed at one level of complexity simply does not exist at a lower level.

Checkland's discussion of complexity necessarily raises the following question, which hitherto has not been asked: where in the hierarchy of levels of complexity, and hence the levels of science, do problems of human performance in systems occur? By and large, the early human factors pioneers were practitioners of the science of psychology. They saw the wartime problem of the increased complexity of man-machine interactions as a problem of psychology and hence, of the particular level of phenomenon dealt with by that science. Being steeped in the methods of science, they applied reductionist thinking to the problem, dividing it into concerns of selection, training, and equipment design. As suggested in the lectures by Chapanis and colleagues (1947), they were conscious of the fact that each of these avenues of inquiry was but an aspect of the total problem of addressing human performance in systems. Nevertheless, they likely felt justified in taking a reductionist approach since the doctrine of dividing physical phenomena into separate parts was an unquestioned part of the scientific perspective of their world. In so doing, they necessarily accepted the fundamental assumption that such division did not distort the phenomena they were studying. However, nearly 25 years

later, the work by Simon (1976), reviewing the prior two decades of human factors experimental results, challenged the veracity of this assumption:

> Generalizable experimental data that will predict performance quantitatively and with reasonable accuracy is not likely to be generated from experiments that examine only a few factors. The world is more complex than any two-, three-, or four-factor study is likely to approximate. More factors must be examined before predictive precision can be achieved…(p. 92).

It thus appears that the human factors sciences had become unduly restricted as a result of discarding from the outset a great deal of the complexity inherent in the problem of human performance in systems. By selecting simple subsets of the problem for examination and controlling others, they introduced a systematic bias into any picture of human performance that was based on them. Indeed, there is the possibility that the human factors sciences, based upon reductionism, repeatability, and refutation, foundered when faced with extremely complex phenomena that entailed more interacting variables than they could cope with in their experiments:

> Traditional ergonomics has failed to significantly improve overall *system* [emphasis in original] productivity, worker health, and the intrinsic motivational aspects of work systems…progressively more examples were being seen where organizational systems with good traditional micro-ergonomic design were not achieving overall organizational goals because of a failure to address the macro-ergonomic design of the work system (Hendricks, 1995, p. 1618).

Overall then, the first three decades of the post-war period demonstrated that the human factors sciences could provide an accounting of the mechanisms of some human performance phenomena, but they could not fully explain the problem of human performance in systems. In other words, the problem of human performance in systems could not be effectively collapsed into the human factors sciences. There was some aspect of the problem that emerged at higher levels of complexity that simply could not be resolved through reductionist thinking—that is, irreducible complexity.

Not surprisingly, the 1980s saw the emergence of two "large-systems" disciplines, macroergonomics[1] and HSI (Table II-1), both of which are founded, to varying degrees, on sociotechnical systems theory (Kleiner, 2008). Sociotechnical systems theory, in turn, views organizations as open systems engaged in the process of transforming inputs into desired outcomes. They are open because the work system boundaries are permeable and exposed to the environment in which they function and on which they depend. Organizations bring two critical factors to bear on the transformation process: technology in the form of a *technological subsystem*, and people in the form of a *personnel subsystem*. The design of the technological subsystem primarily defines the tasks to be performed, whereas the design of the personnel subsystem prescribes the ways in which the tasks are performed. The two subsystems interact with each other, are interdependent, and operate under the concept of *joint causation*, meaning that both subsystems are affected by causal events in the environment. The technological subsystem, once designed, is fixed and whatever adaptation the organization permits falls to the personnel subsystem to implement. Joint causation underlies a related key sociotechnical systems concept, namely *joint optimization*. Since the technological and personnel subsystems respond jointly to causal events, optimizing one subsystem and then fitting the second to it results in suboptimization of the joint work system. Consequently, joint optimization requires the integrated design of the two subsystems to develop the best possible fit between the two given the objectives and requirements of each and the overall work system (Meister, 1999; Hendrick & Kleiner, 2002).

[1] Industrial/organizational (I/O) psychology, which includes the field of organizational behavior, predates and is distinguished from macroergonomics (Muchinsky, 1993). I/O psychology is primarily concerned with selecting people to fit work systems, in contrast to ergonomics and human factors engineering, which focus on designing work systems to fit people. In turn, macroergonomics can be viewed as the opposite side of the coin from organizational psychology. Both organizational psychology and macroergonomics are concerned with the design of organizational structures and processes, but their focus is somewhat different. Common objectives of organizational psychology include improving motivation and job satisfaction, developing effective incentive systems, enhancing leadership and organizational climate, and fostering teamwork. While these objectives also are important to macroergonomics, the primary focus of macroergonomics is to design work systems that are compatible with an organization's sociotechnical system characteristics; and then to ensure that the micro-ergonomic elements are designed to harmonize with the overall work system structure and processes (Hendrick & Kleiner, 2002).

Table II-1. Comparison among three large-system approaches [After Kleiner, 2008].

Approach	Macroergonomics	Human systems integration	Systems engineering
Theory	Sociotechnical systems theory	Loosely on sociotechnical systems theory	Systems theory
Primary level of focus	Work-system/organization	Technological systems, subsystems, and small systems/devices	Systems
Additional foci	Environment, personnel, technology	Functions of human factors, manpower, personnel, training, systems safety, health hazards, and survivability	Integration of functions
Primary performance impact	Productivity, health and safety, satisfaction, culture	System performance and life-cycle cost	System
Additional value-added characteristics	Macro-to-micro linkage, especially applicable to human factors professionals	Especially applicable to military systems	Integrates technical functions

Checkland (1981) describes the systems movement as an effort to investigate the implications of using the concept of the irreducible whole, or "a system," in any area of endeavor. He asserts that the systems movement is not itself a discipline, but a way of thinking about problems that can be applied to any of the arbitrary divisions of human knowledge known as disciplines. He explains systems thinking as "an attempt, within the broad sweep of science, to retain much of that tradition but to supplement it by tackling the problem of irreducible complexity via a form of thinking based on wholes and their properties which complements scientific reductionism" (p. 74). Within this context, it is easy to see that sociotechnical systems theory ascribes to a systems approach: it focuses on an emergent property of a whole, namely the ability of an entity to transform inputs

into desired outcomes, which results from the degree of optimization of the integration of its technological and personnel subsystems. Those disciplines that have arisen from sociotechnical systems theory, such as macro-ergonomics, HSI, systems ergonomics and human factors integration in the United Kingdom, and perhaps even the total quality management movement in the United States, all share a "large-system perspective" in which the primary level of focus is a system (Kleiner, 2008). Hence, one would expect to find systems thinking scientists, engineers, technologists, psychologists, and management scientists within these disciplines.

B. THE TRANSITION FROM MACHINE AGE TO SYSTEMS AGE

As previously discussed, the pioneering behavioral scientists of World War II, typified by Chapanis and colleagues, used the rational approach to epistemology known as reductionism to deal with the complexity inherent in the problem of human performance in systems. This philosophy of reductionism is often attributed to Descartes, as presented in his *Discourse on the Method* (1637), and is based on four precepts:

- Accept as true only what is definite
- Divide every question into manageable parts
- Begin with the simplest issues and ascend to the more complex
- Review frequently enough to retain the whole argument at once.

Reductionism gives rise to the analytical way of thinking whereby understanding of the world is the sum, or result, of an understanding of its parts. Reductionist analysis involves decomposing phenomena into independent and indivisible parts, explaining the behavior of these individual parts, and then aggregating these partial explanations into the explanation of the whole. All phenomena are explainable by mechanisms consisting of one simple relationship: cause and effect. The cause is both necessary and sufficient for the effect. The prevailing view of the world is then deterministic and there is no need for teleological concepts. This approach predisposes one to thinking of the world as a machine, which led Russell Ackoff (1981) to refer to reductionist thinking as "Machine Age." Ackoff claims we are now in the "Systems Age," which necessitates a different approach (Table II-2).

Table II-2. Machine Age versus Systems Age paradigms [From Ackoff, 1981].

Machine Age procedure	Systems Age procedure
Decompose that which is to be explained Explain the behavior or properties of the contained parts separately Aggregate these explanations into an explanation of the whole	Identify a containing system of which the thing to be explained is a part Explain the behavior or properties of the containing whole Explain the behavior of the thing to be explained in terms of its role(s) and function(s) within its containing whole
Machine Age analysis	**Systems Age synthesis**
Analysis focuses on structure; it reveals how things work	Synthesis focuses on function; it reveals why things operate as they do
Analysis yields *knowledge*	Synthesis yields *understanding*
Analysis enables description	Synthesis enables explanation
Analysis looks *into* things	Synthesis looks *outward* from things

In contrast to Descartes' philosophy of reductionism, Aristotle argued that the whole was more than the sum of the parts, and the form of the whole signified its function, and hence, its intrinsic purpose or *telos*. Although Aristotle's doctrine was superseded by the Scientific Revolution of the 17th century, his ideas have, in part, been reinstated in the Systems Age concepts of expansionism and teleonomy, which replace reductionism and cause-and-effect relationships. Expansionism is a doctrine maintaining that all objects and events, and all experiences of them, are parts of larger wholes; teleonomy is a doctrine in which structures and behaviors are determined by the purpose they fulfill. Synthesis is simply the combination of parts so as to form a whole, and thus, synthetic thinking requires integrating things within a containing, or parent, system and explaining them in terms of their role(s) and function(s) in the parent system. Expansionism requires synthetic thinking, whereby attention is turned from ultimate elements to the whole with interrelated parts—"systems." Phenomena are explained in terms of probabilistic producer-product relationships such that a producer is only necessary but not sufficient for its product, and by implication, cannot provide a complete

explanation of it. The prevailing view of the world is then stochastic and there is a need to look at systems teleonomically, in an output- or outcome-oriented way, rather than deterministically, in an input-oriented way (Ackoff, 1981; Checkland, 1981).

Returning to our problem of interest of explaining human performance in a system, a Machine Age thinker would likely start by first considering the system functions allocated to the human. They might then identify the tasks that must be performed in accomplishing those functions, assess the aptitude profile of the prototypical user (i.e., personnel selection), examine the human-machine interface (i.e., human factors engineering), and review the training curriculum. Finally, the Machine Age thinker would aggregate these considerations to explain how the human should perform in the system. In contrast, a Systems Age thinker would start by identifying a system containing the human, say the personnel system, and would then define the functions or objectives of the personnel system with reference to an even wider social system that contains it, such as an organization or military service. Finally, they would explain the roles or functions of the human in the system with reference to the objectives of the personnel system. While it is clear that analysis (i.e., reduction) and synthesis (i.e., expansion) both have a role when considering human performance in systems, it is synthesis that is the particular focus of sociotechnical systems theory.

C. GESTALT AND HOLISM

The word *Gestalt*, while not having an exact equivalent in English, is used in German to mean the way a thing has been "placed" or "put together." In psychology, the word is often implied to mean "pattern" or "configuration." Gestalt philosophy emerged in the early twentieth century, mainly in Germany and Austria, in reaction to reductionism. The field of Gestalt psychology was launched in 1912 with the publication of Max Wertheimer's study of the visual illusion of movement resulting from the serial presentation of still images. Gestalt psychology was based on the observation that we often experience things that are not a part of our simple sensations. Gestalt theory was holistic and embraced the concept of emergent properties:

> The fundamental "formula" of Gestalt theory might be expressed in this way: There are wholes, the behavior of which is not determined by that of their individual elements, but where the part-processes are themselves determined by the intrinsic nature of the whole. It is the hope of Gestalt theory to determine the nature of the wholes (Wertheimer, 1938, p. 2).

While early Gestalt work was primarily concerned with perception, particularly visual perception of illusions, the approach was later extended to problems in other areas such as social psychology and economic and political behavior.

Despite the fact that von Bertalanffy (1972), a founder of general systems theory, mentions Gestalt theory as a historical prelude, Hitchins (1992) asserts that the legacy of Gestalt theory is often overlooked despite its central tenets being deeply embedded in modern systems thinking. Contemporary HSI philosophy, for example, with its focus on the synthesis of human performance in systems seems to owe as much to Gestalt theory as to the original human factors sciences. For instance, the Gestalt principle of *reification* addresses the constructive aspect of perception whereby the experienced entity contains more spatial information than its component sensory stimuli. Illustrated in Figure II-2A, the component sensory stimuli consist of three black wedges, but the viewer likely perceives a triangle, with the wedges at the vertices, even though no triangle has actually been drawn. The triangle is an emergent property of the whole that cannot be discerned when the component stimuli are examined out of context from the whole (Figure II-2B).

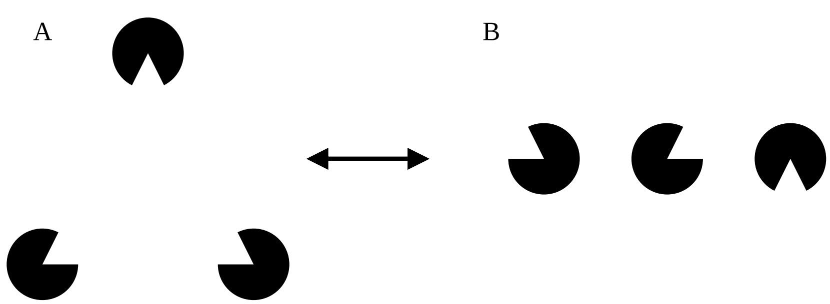


Figure II-2. Illustration of the Gestalt principle of reification.

By analogy, human performance is an emergent property of the whole of individual abilities (i.e., personnel domain), training, and equipment design (i.e., human factors engineering), among other determinants. Likewise, total system performance is an emergent property of the whole of the personnel and technological subsystems. Human performance in systems cannot be fully discerned through the independent examination of foci like human factors, manpower, personnel, training, system safety, health hazards, etc. The concepts of holism and emergence could well explain Simon's (1976) observation after reviewing 239 human factors engineering experiments: between one-third and one-fifth of the variance in human performance is not attributable to an interpretable source.

D. EMERGENCE AND HIERARCHY

We just demonstrated the concept of emergence in the context of a simple system of symbols, and we will now expand this concept to hierarchical systems in general. Figure II-3 illustrates *n* nested systems such that each system contains, and is contained, by other systems.

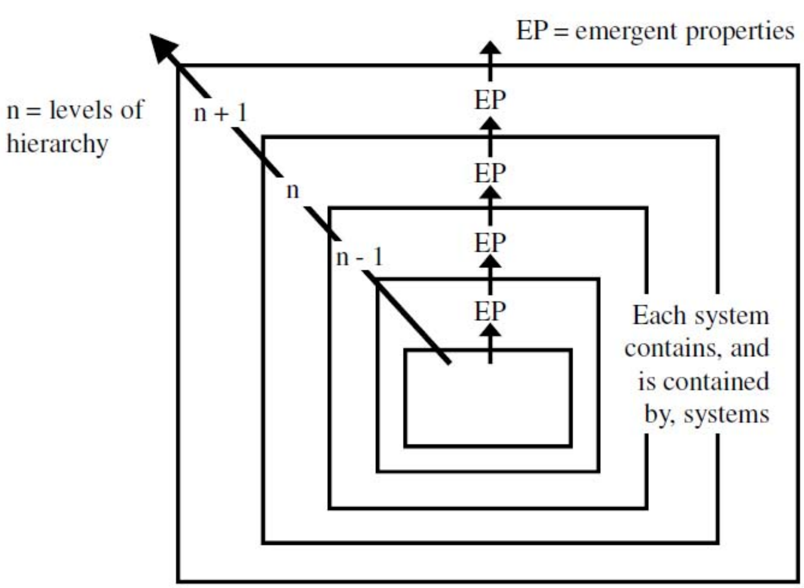


Figure II-3. Emergence and hierarchy [From Hitchins, 1992].

The *principle of emergence* states that whole entities exhibit properties that are meaningful only when attributed to the whole and not the parts. For example, the human brain exhibits self-awareness, but this property cannot be attributed to any specific locus within the brain. Every system exhibits emergent properties that derive from its component activities, interactions, and structure but which cannot be reduced to the individual components. Often it is these emergent properties that make a system purposeful or yield value to system stakeholders (Hitchins, 1992).

According to the *principle of hierarchy,* entities that can be meaningfully considered as wholes are built up from smaller entities, which themselves are whole, and so on (Hitchins, 1992). For example, the human is composed of several systems such as the digestive, cardiovascular, neurological, reproductive, and musculoskeletal systems, to name just a few. The cardiovascular system, in turn, is composed of the heart, arteries, veins, capillaries, etc. Continuing the example, but instead ascending in the hierarchy, it also may be observed that individual humans are components of larger social systems such as teams, families, or communities.

These concepts of emergence and hierarchy are fundamental to the systems movement and systems thinking (Checkland, 1981). In a hierarchy, emergent properties correspond to levels. The key insight, then, is that a system-of-interest can only be meaningfully observed from the level of its containing system (Hitchins, 1992). For example, in Figure II-3, if one were interested in the $(n-1)^{\text{st}}$ system, it would be necessary to observe it from the n^{th} level to perceive its emergent properties. Checkland (1981) goes so far as to explain emergence and hierarchy in terms of a generalized model of organized complexity:

> …the general model of organized complexity is that there exists a hierarchy of levels of organization, each more complex than the one below, a level being characterized by emergent properties which do not exist at the lower level. Indeed, more than the fact that they 'do not exist' at the lower level, emergent properties are *meaningless* [emphasis in original] in the language appropriate at the lower level. 'The shape of an apple,' although the result of processes which operate at the level of the

> cells, organelles, and organic molecules which comprise apple trees…*has no meaning* at the lower levels of description. The processes at those levels result in an outcome which [sic] signals the existence of a new stable level of complexity—that of the whole apple itself—which has emergent properties, one of them being an apple's shape (p. 78).

Given Checkland's perspective, we might reasonably ask ourselves whether human performance in systems "has any meaning" at the lower levels of description created by the human factors sciences and corresponding HSI domains. This might explain, for instance, why human factors engineers choose to sidestep the topic:

> If the system hierarchy breaks down into the workstation, the subsystem, and the total system, the systems concept requires measurement at all these levels and the determination of the relationships among them. Again, this is a conceptual requirement that is usually ignored (Meister, 1999, pp. 144–145).

Meister appears to suggest that the human factors community has a preference to avoid traversing levels of organized complexity.

While emergence and hierarchy are not yet phrases in general use among HSI practitioners, they need to be thought of as twin components of any HSI philosophy. For example, it is both reasonable and insightful to describe the primary task of HSI in terms of emergence as:

> Integrating humans with other system elements to form and maintain a system with the requisite emergent properties to meet stakeholders' needs.

It is equally constructive to consider where in the system hierarchy HSI is appropriately addressed. For instance, the HSI domains of human factors engineering, manpower, personnel, training, system safety, etc., can meaningfully be considered as individual systems, both in the sense of organizations within the Defense Department and as bodies of knowledge. To consider the emergent properties of the synthesis of these systems, HSI must reside at the level of their containing system. So, as a philosophy, HSI is best considered a meta-discipline, sitting "above" the constituent domains, seeking to provide an umbrella over them, and establishing a comprehensive set of unifying perspectives. Thus, it is the principle of hierarchy that distinguishes HSI from the corresponding field of human factors. Likewise, from the perspective of sociotechnical systems theory, joint

optimization is an emergent property that must be considered from the level of the system that contains both the personnel and technological subsystems. Consequently, HSI as a program should be implemented at a level in the Defense Department that has oversight and/or coordination responsibility for both of these subsystems.

E. SYSTEMS TYPOLOGIES

A major premise of the systems movement is the notion that it is more insightful to view the apparently chaotic universe as being comprised of a complex of interacting wholes called "systems" rather than as a set of phenomena whose laws can be established by the reductionist experimental approach. Given this hypothesis, it is not surprising, then, that a number of general attempts have been made to describe and classify the possible types of systems (Checkland, 1981). Kenneth Boulding (1956), a founding father of general systems theory, proposed one of the first general classifications of systems types (Table II-3).

Table II-3. Boulding's classification of systems [After Boulding, 1956].

Level	Characteristics	Examples	Relevant disciplines
1. Structures, frameworks	Static	Crystal structures, bridges	Description, verbal or pictorial, in any discipline
2. Clock-works	Predetermined motion (may exhibit equilibrium)	Clocks, machines, the solar system	Physics, classical natural science
3. Control mechanisms	Closed-loop control	Thermostats, homeostasis mechanisms in organisms	Control theory, cybernetics
4. Open systems	Structurally self-maintaining	Flames, biological cells	Theory of metabolism (information theory)
5. Lower organisms	Organized whole with functional parts, "blue-printed" growth, reproduction	Plants	Botany

Level	Characteristics	Examples	Relevant disciplines
6. Animals	A brain to guide total behavior, ability to learn	Birds and beasts	Zoology
7. Humans	Self-consciousness, knowledge of knowledge, symbolic language	Human beings	Biology, psychology
8. Socio-cultural systems	Roles, communication, transmission of values	Families, social groups, organizations, nations	History, sociology, anthropology, behavioral science
9. Transcendental systems	"Inescapable unknowables"	The idea of God	?

Boulding sought to organize "individual" units as found in empirical studies of the real world into an informal, intuitive hierarchy based on their relative degree of complexity. Within this hierarchy:

- Emergent properties are assumed to arise at each defined level
- Complexity increases as one ascends the hierarchy
- Lower-level systems, and their distinguishing properties, are found in higher-level systems.

Boulding's objective, given the emergence at the time of an increasing number of hybrid disciplines, was to provide a framework of complexity within which one could relate the different empirical sciences. Accordingly, we can view the historical development of HSI, itself a hybrid discipline, as an attempt to bring in HSI domain-related disciplines to treat problems at levels 7 and 8.

Nehemiah Jordan (1968) proposed a second general systems taxonomy based on three organizing principles that he asserts allow us to perceive a group of entities as a proper system. These principles include rate of change, purpose, and connectivity, and each is defined in terms of a pair of systems properties that are polar opposites (Figure II-4).

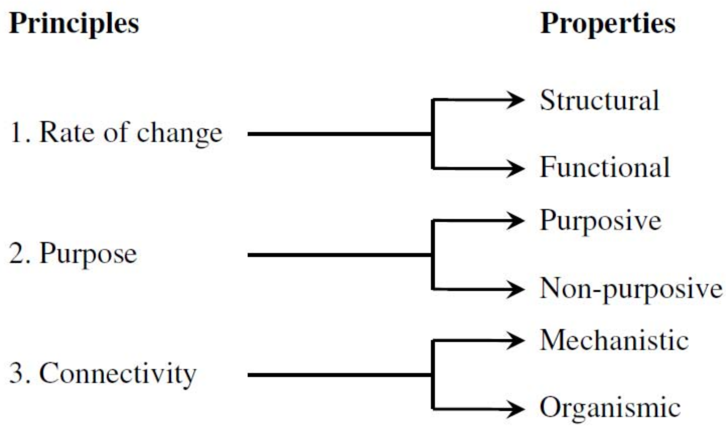


Figure II-4. Jordan's taxonomy [After Jordan, 1968].

Thus, a system is characterized as *structural* if the rate of change is slow or *functional* if the rate of change is fast. Systems are either *purposive* or *non-purposive*. They are *mechanistic* if the parts of a system are not strongly interdependent, or they are *organismic* if such interdependence is strong. Jordan argues that we should only use "dimensional" descriptions, his principles and properties being prototypes, when talking about systems. Hence, there are 2^3 or eight ways of selecting one from each of his three pairs of properties to form potential descriptions of groupings worthy of the name "system." For example, a system to carry out HSI might be described as having a functional, purposive, and organismic system of domains (i.e., manpower, personnel, training, system safety, etc).

Checkland (1981) also proposed a systems typology, or systems map of the universe, consisting of five classes of systems (Figure II-5).

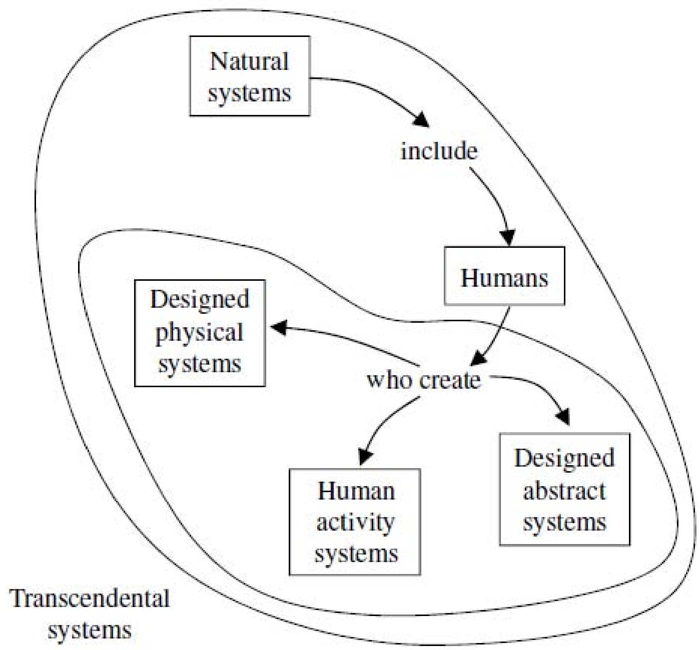


Figure II-5. Checkland's systems map of the universe [From Checkland, 1981].

Natural systems are the physical systems that make up the universe, to include the living things observed on Earth. These can be considered as "given" systems because their origin is that of the universe and they result from the processes of evolution. There are many other systems that are similar to natural systems, with the very important exception that they result from human conscious design. Unlike natural systems, such systems could be made to be other than they are. These include *designed physical systems*, which range from simple hand tools to spacecraft, and *designed abstract systems* such as mathematics and philosophies. These systems are brought into existence to serve some human purpose, although at times that purpose may be hard to define explicitly. Then there is the human act of design, which itself is a fourth possible system class, namely the *human activity system*. Human activity systems are less tangible than natural and designed systems, but nevertheless, are clearly observable in the world as sets of human activities more or less consciously ordered into wholes as a result of some underlying purpose or mission. This is a very broad class of systems, ranging from the extremes of a single artist wielding a paintbrush to international political systems

working to make life more tolerable for the human race. Beyond these four classes of systems, there is a category called *transcendental systems*, which includes those systems that are beyond knowledge (e.g., God). The systems map suggests that the absolute minimum number of systems needed to describe the whole of reality is four: natural, designed physical, designed abstract, and human activity systems.

Following Checkland's paradigm, HSI is concerned with the combination of natural systems, in the form of humans, and designed systems (more often than not designed physical systems) to form wider systems showing emergent properties that are coherent with the purpose or mission of at least one human activity system. From the broader perspective of sociotechnical systems theory and the related concept of joint optimization, HSI may also be described as being concerned with the combination of human activity systems (i.e., the personnel subsystem) and designed physical systems (i.e., the technological subsystem). When the idea of HSI emerged as a conscious product of the human mind, it was a designed abstract system. When the idea was captured and translated into text in the form of published articles, books, and Defense Department policy guidance, it became a designed physical system. As an organizational activity, HSI is itself the purpose or mission of specific human activity systems. While perhaps confusing, this should reinforce the importance of perspective, or *Weltanschauung*, in systems thinking. You must know the perspective from which observations are made. This distinction separates reductionist studies of natural systems, which can be made independent of the perspective of the observer, and hence the definition of scientific fact, from that of human-created systems (Checkland, 1981). While few would dispute that a human is a natural system, such certainty is not the case when we describe "HSI." The latter does not fit easily into any one system class, and so it is not easy to obtain descriptions of HSI upon which all observers can agree (Deal, 2007).

F. COMPLEX ADAPTIVE SYSTEMS

Many human and social systems can be likened to complex adaptive systems, a notion derived from the study of non-equilibrated natural systems such as physical,

chemical, and biological systems (Holland, 1995). Classical science, based on the reductionist view of the world, considers entities as independent and treats systems as being close to equilibrium. System dynamics are considered to be linear; models and theories are validated if they can accurately predict experimental results. Complexity science recognizes that entities are interdependent and many systems studied are far from equilibrium, giving rise to non-linear system dynamics. Complex systems exhibit self-organization, emergence, and evolution; understanding is no longer demonstrated by prediction, but rather by an awareness of the limits of predictability (Holland, 1995, 1998; Lyons, 2004).

As an illustration, consider Figure II-6 depicting a human activity system. The entities in the human activity system are the component teams performing specific activities supporting a common purpose or mission. These teams interact and connect with each other in unpredictable and unplanned ways. Certain regularities emerge from the mass of their interactions and form a pattern that feeds back to the human activity system and informs the interactions of its constituent teams.

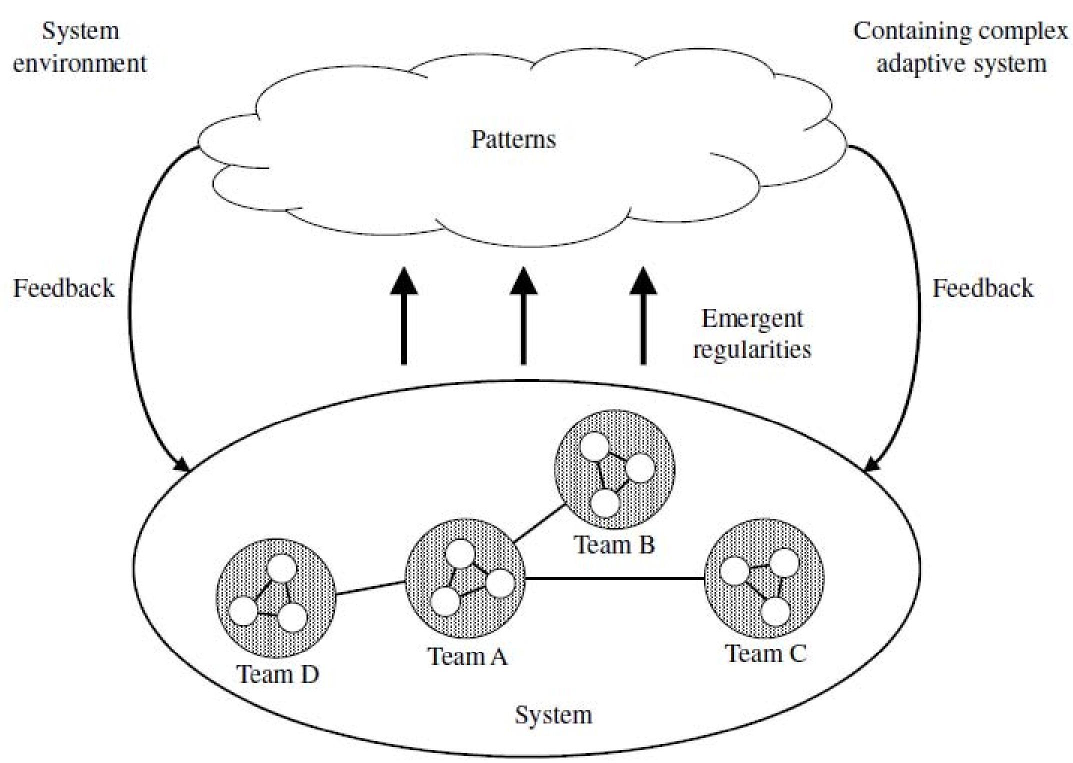


Figure II-6. Complex adaptive systems.

Now suppose the teams comprising our system-of-interest are each functionally specialized such that the members of any one team are not directly interchangeable with those on other teams. Let us assume changing population demographics or economic factors deplete the number of individuals that can be recruited and trained to staff one of the teams. For example, a change occurs in the personnel system that contains our system-of-interest and hence contributes to its external environment. This change may result in either more or less work for other teams in the system and will affect their behavior and size. A period of flux occurs in all the teams in the system until a new balance is established. It is reasonable to expect diminished system performance during the acute period of heightened system disequilibrium. However, system adaptation may also result in chronic performance decrements, increased system losses, and/or excessive ownership costs. Aptly, this example then illustrates a typical HSI challenge: to anticipate and/or manage system adaptations, thereby increasing the likelihood that designed systems meet their stated purpose. At a macro-level, it also captures the very scenario that led to the development of a Defense Department HSI program (which was the problem of economically recruiting personnel of sufficient quality to match the influx of substantial amounts of technologically advanced equipment into the military services) (Booher, 1990).

Before moving on, it is worth briefly discussing chaos since the concept is often mentioned in connection with complex adaptive systems. Complexity theory is distinct from chaos theory, but the idea of chaos still plays an important role in complexity theory. Systems can be considered as existing along a spectrum ranging from equilibrium to chaos. A system in equilibrium lacks the internal dynamics to enable it to respond to its environment and it will subsequently die. At the other extreme, a system in chaos ceases to function as a system and dies as well. The most productive state then for a system is to be “on the edge of chaos” where there is maximum internal dynamics, and hence capability to respond to the environment (Lyons, 2004).

G. WICKED PROBLEMS

Systems thinking is a general intellectual framework for approaching problem situations. Pidd (2003) discusses the ways in which people use the term "problem" and provides a spectrum containing three points as examples:

- *Puzzles:* Situations with clear objectives and where solutions are obtained by applying known methods
- *Problems:* Well defined and structured situations, but requiring considerable ingenuity and expertise to solve
- *Messes:* Ill-defined and unstructured situations for which there is considerable disagreement over objectives; must be structured and shaped before any solution, should such exist, can be found.

Rittel and Webber (1973), working on problems of policy planning, describe Pidd's messes as "wicked problems." Their choice of the term "wicked" is not meant to characterize certain problems as having properties that are ethically deplorable, but rather to assert that they are "vicious" or "tricky." Problems in the natural sciences, such as those worked by scientists and some classes of engineers, have both clear objectives and identifiable "optimal" solutions making them "tame." Wicked problems, in contrast, have neither of these clarifying traits. Figure II-7 considers puzzles, problems, and messes/wicked problems within the spectrum of systems modeling approaches. At one extreme, models are used to totally or partially replace humans in routine decision making such as by providing automated decision making or routine decision support. At the other extreme, models are used to support people who are thinking through difficult issues, either by representing possible system designs or by representing insights that are debated (Pidd, 2004).

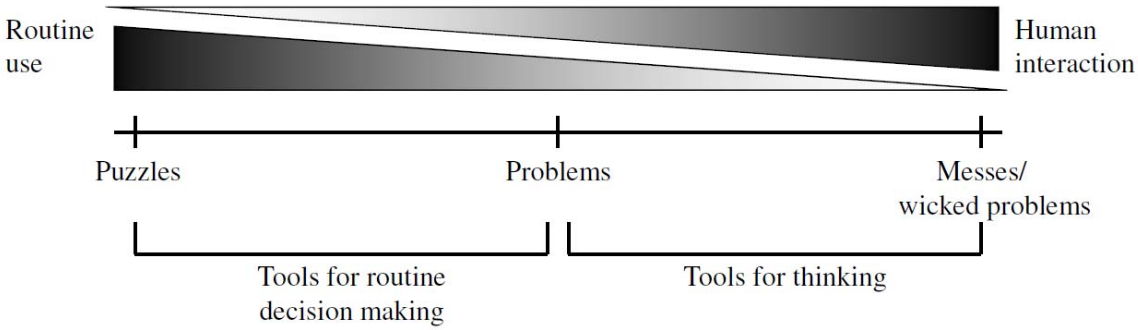


Figure II-7. Spectrum of problems with corresponding systems modeling approaches [From Pidd, 2004].

Rittel and Webber (1973) propose ten distinguishing properties of wicked problems for which systems practitioners should be alert:

- There is no definitive formulation of a wicked problem. The information needed to *understand* the problem depends upon one's ideas for *solving* it.
- Wicked problems have no stopping rule. Work stops on the problem not for reasons inherent to the logic of the problem, but for considerations that are external to the problem such as time or funds.
- Solutions to wicked problems are not true-or-false, but good-or-bad. There are no conventionalized criteria for objectively deciding whether a solution is correct, only relative judgments of "goodness."
- There is no immediate and no ultimate test of a solution to a wicked problem. Solutions to wicked problems will generate waves of consequences over an extended, perhaps even virtually unbounded, period of time.
- Every solution to a wicked problem is a "one-shot operation" because there is no opportunity to learn by trial-and-error; every attempt counts significantly. Any solution actions are effectively irreversible and the half-lives of their consequence are very long.
- Wicked problems do not have an enumerable set of potential solutions, nor is there a well-described set of permissible operations that may be incorporated into plans.

- Every wicked problem is essentially unique. "Essentially unique" means that the current problem, while sharing many similarities with previous ones, has a distinguishing property that is of overriding importance.
- Every wicked problem can be considered a symptom of another problem. This characteristic implies a hierarchy of problems; consequently, the challenge becomes to determine the appropriate level for intervention.
- The existence of a discrepancy representing a wicked problem can be explained in numerous ways, and the choice of explanation determines the nature of the problem's resolution (e.g., crime can be explained by not enough police, too many guns, or insufficient socioeconomic opportunities, each of which offers a different approach to attacking crime). An analyst's *Weltanschauung* is the strongest determining factor in explaining a discrepancy, and therefore, in resolving a wicked problem.
- The aim is not to find truth but to improve some characteristics of the world in which people live.

HSI problems, which involve the challenge of bringing together human activity systems and designed physical systems, often exhibit attributes of wicked problems. For example, the Defense Department objectives for HSI include both optimizing total system performance and minimizing total ownership costs (DoD, 2008). Since performance and cost objectives are usually inversely correlated, progress towards either objective will be at the expense of the other. Given the absence of guidance on relative priority, these two objectives are often assumed to be equally important. One would have to conclude then that the objective of the Defense Department's HSI program is constantly to maintain and adjust a politically acceptable balance between these incompatible objectives. These types of problems are distinctly different from human factors problems that can be technically defined: e.g., "reduce assembly mean time to repair to 8 hours," or "the interface must accommodate the 5^{th} percentile female." While the methods of the human factors disciplines play a prominent role in HSI, they become operational only after the most important decisions have been made – after the wicked planning problem has already been tamed.

H. SOFT AND HARD SYSTEMS APPROACHES

Pure hard and soft systems approaches represent extreme points on a spectrum, and Table II-4 presents the way theorists and practitioners view the differences (Pidd, 2004). To understand the two archetypal systems approaches, it is helpful to distinguish between two extreme types of rationality (Simon, 1982). The first type, *substantive rationality*, is what most people assume when discussing rational analysis. It is based on the notion that:

- A set of alternative courses of action can be presented for an individual's choice
- Data and information are available that permit the individual to predict the consequences of choosing any alternative
- A criterion exists for determining the preferred set of consequences.

The course of action that leads to the most preferred set of consequences is then selected by the individual. When problem situations recur, many mathematical and statistical models can be used to help manage situations that meet the requirements specified above.

Table II-4. Practical aspects of hard and soft systems approaches [From Pidd, 2004].

	Hard approaches	Soft approaches
Methodology	Based on common sense, taken-for-granted views of analysis and intervention	Based on rigorous epistemology
Models	Shared representations of the real world	Representations of concepts relevant to the real world
Validity	Repeatable and comparable with the real world in some sense	Defensibly coherent, logically consistent, plausible
Data	From a source that is defensibly there in the world with an agreed or shared meaning, observer independent	Based on judgment, opinion, some ambiguity, observer-dependent
Values and outcome of the study	Quantification assumed to be possible and desirable. From option comparison based on rational choice	Agreement (on action?), shared perceptions. Informing action and learning
Purpose	For the study: taken as a given at the start	For the study: remains problematical
	For the model: understanding or changing the world, linked to the purpose	For the model: a means to support learning

In contrast, a second type of rationality, *procedural rationality*, is applied in situations that are novel and irregular as in the case of wicked problems. Procedural rationality stresses processes to support decision making based on human deliberation when substantive rationality is impossible or impracticable. Procedural rationality is based on the following notions:

- Options or courses of action must be discovered
- Acceptable solutions must be developed by resolving conflict over ends and means
- Information and analysis are still crucial but are bounded by cognitive and economic limitations

- Individuals tend to satisfice across known acceptable solutions rather than work to discover globally optimal solutions.

HSI problem situations, more often than not, require the application of procedural rationality, thereby lending them to analysis using the “softer” approaches, particularly in the early stages of the problem situation. For example, as previously mentioned, the Defense Department’s objectives for its HSI program include both optimizing total system performance and minimizing total ownership costs (DoD, 2008), but no description or definition of “optimization” is provided in the program guidance. Consequently, there exists no *a priori* quantitative criterion for determining the preferred set of consequences vis-à-vis HSI as is required to apply substantive rationality and make use of the hard systems approaches. This point was aptly demonstrated during a recently developed graduate level course on HSI at the Naval Postgraduate School in which students were asked to formulate a key performance parameter (i.e., a single quantitative criterion) for HSI. A survey of the students’ responses revealed that almost none of them were able to provide a relatively tractable or robust criterion. However, the larger issue for HSI as a discipline is that it is a hybrid, in part, of the social sciences—recall HSI’s lineage in sociotechnical systems theory! As described by Checkland (1981), the social sciences pose a particular challenge because they are “unrestricted” sciences:

> In a restricted science such as physics or chemistry a limited range of phenomena are studied, well-designed reductionist experiments in the laboratory are possible, and it is probable that far-reaching hypotheses, expressed mathematically, can be tested by quantitative measurements…In an unrestricted science…the effects under study are so complex that designed experiments with controls are often not possible. Quantitative models are more vulnerable and the chance of unknown factors dominating the observations is much greater (p. 65).

Any “unrestricted” science in Checkland’s sense will present considerable problems for those seeking to employ the hard systems approaches.

Although the hard and soft approaches in systems modeling are often presented as polar opposites, there is a growing interest among the corresponding disciplines in looking at the combined use of the two approaches—something Pidd (2004) calls “complementarity in systems modeling.” Pidd suggests three possible relationships

between the hard and soft systems approaches which are represented in Figure II-8. In the left-hand part of the figure, the soft and hard systems approaches are shown as being completely distinct and incommensurable, while in the middle part of the figure, the two are seen feeding off one another in a pragmatic way. In the right-hand part of the figure, the soft systems approaches are shown as containing the classical hard systems approaches, implying that the understanding gained from the soft systems approaches enables a sensible attempt at the hard approaches. Pidd does not explicitly endorse any of these perspectives, but he does offer Flood and Jackson's (1991) Total System Intervention as a prototype approach for achieving complementarity.

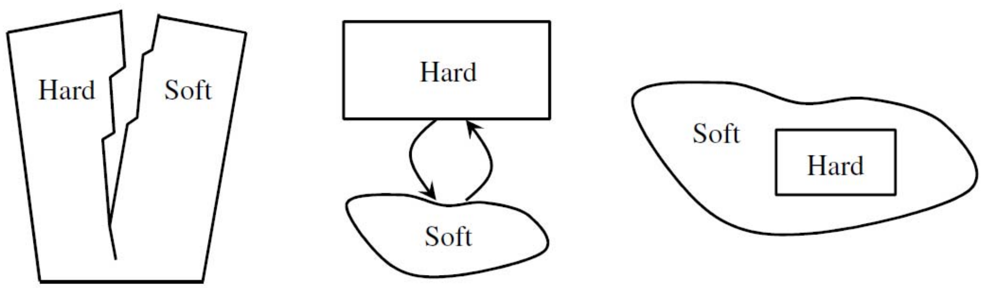


Figure II-8. Relationships between hard and soft systems approaches [From Pidd, 2004].

I. TOTAL SYSTEMS INTERVENTION

The existence of the hard and soft system dichotomy appears to break with the original holistic intent of the systems perspective. Flood and Jackson (1991) attempt to remedy this situation with their Total Systems Intervention (TSI) approach to problem solving, which demonstrates that all problem solving methods can be arranged and operated successfully as an organized whole. In essence, TSI serves as a meta-methodology that enables problem solvers to employ a variety of methods by first creatively surfacing issues an organization faces and then choosing the method(s) best equipped to tackle those issues most effectively (Flood, 1995). TSI is constructed on a theoretical foundation called Critical Systems Thinking, which assumes the following three positions (Flood & Jackson, 1991):

- *Complementarism*: The existence of a range of systems methodologies, each driven by a different theoretical position, is a strength of the systems movement if each methodology is put to work only on the kinds of problems for which it is the most suitable.
- *Sociological awareness*: Organizational and societal pressures exist that lead at times to certain systems methodologies being popular for guiding interventions, making it necessary for problem solvers to contemplate the social consequences of using a particular methodology.
- *Human well-being and emancipation*: The exercise of power in the social process can limit the open and free discussion necessary for successful mutual understanding among all those involved in social systems; human beings have an "emancipatory interest" in freeing themselves from constraints imposed by power relations and in learning to control their own destiny.

TSI can be studied through its philosophy, principles, and process. The philosophy describes the worldview from a TSI perspective, the principles propose the kinds of action that should be taken, and the process sets out how to implement the principles.

The TSI philosophy ascribes to the image of a whole that has emergent properties, a hierarchical structure, and processes of communication and control to enable it to adapt and survive in a changing environment. This image provides a framework on which an ideal whole system view of an organization can be constructed in five stages (Flood, 1995):

1) An organization is a horizontally and vertically integrated set of technical and human activities.
2) Activities of an organization must be efficiently and effectively controlled while maintaining viability of the organization.
3) Activities of an organization must be directed to achieve some intended purpose.
4) People in organizations appreciate 1) and 3) in different ways, which can cause conflict, lack of cohesion, inefficiency, ineffectiveness, rigidity, and non-viability in organizations.

5) Both 3) and 4) must be harmonized through organizational design and management style.

Given this view, organizational problem solving equates to managing interrelated sets of issues arising from the interaction of technical and human activities rather than solving identifiable problems. An organization can then be understood in terms of interacting issues, and problem solving may be considered as being part of the continuous process of managing these issues.

There are four principles, or kinds of action, that promote the implementation of the TSI philosophy described above:

1) *Being systemic*: Study the world as if it were systemic, which means taking into account interactions between all technical and human activities at three hierarchical levels (i.e., the system, the subsystems, and the suprasystem) in the process of continuously managing interacting issues.
2) *Achieving meaningful participation*: If we are to develop an adequate appreciation of all interactions between technical and human activities at three hierarchical levels (i.e., the system, the subsystems, and the suprasystem) at any one time, then the perceptions of all people involved and affected must be drawn into the picture.
3) *Being reflective*: Ensure that a whole system understanding is achieved and all issues are acknowledged by reflecting upon the relationship between different organizational interests and where domination over people exists; ensure that all issues are managed using relevant methods by reflecting upon the dominance of favored approaches to intervention.
4) *Striving for human freedom*: Management practices must be based on an explicit ideology that promotes human freedom (i.e., disemprisoning people from dominating structures and decisions so as to encourage open and meaningful debate).

The fourth principle follows from the preceding three principles: human freedom may be achieved through reflection, which in turn helps to achieve meaningful participation, so

as to promote being systemic and taking into account the whole. Hence, taking the whole into account is an important step toward effective problem solving and avoidance of counter-intuitive consequences.

The process of TSI sets out how to implement the four principles, thereby realizing the philosophy. The process is a systemic cycle of inquiry with interaction back and forth between its three phases: creativity, choice, and implementation. As represented in Figure II-9, TSI is a continuous process, with no predetermined start or finish points, which identifies sets of interacting issues and aids their management. Moreover, the process can move in either direction, having both clockwise and counterclockwise modes of operation. In the clockwise mode, the tasks of each of the phases are as follows (Flood & Jackson, 1991; Flood, 1995):

- *Creativity phase*: Identify issues to be dealt with and demonstrate the interacting nature of these issues.
- *Choice phase*: Choose a method(s) that will best manage the interacting issues identified by the creativity phase, tackling the most pressing issues while managing as many issues as possible.
- *Implementation phase*: Employ the chosen method(s) from the choice phase to manage the issues identified by the creativity phase to develop and implement specific change proposals that tackle the given issues.

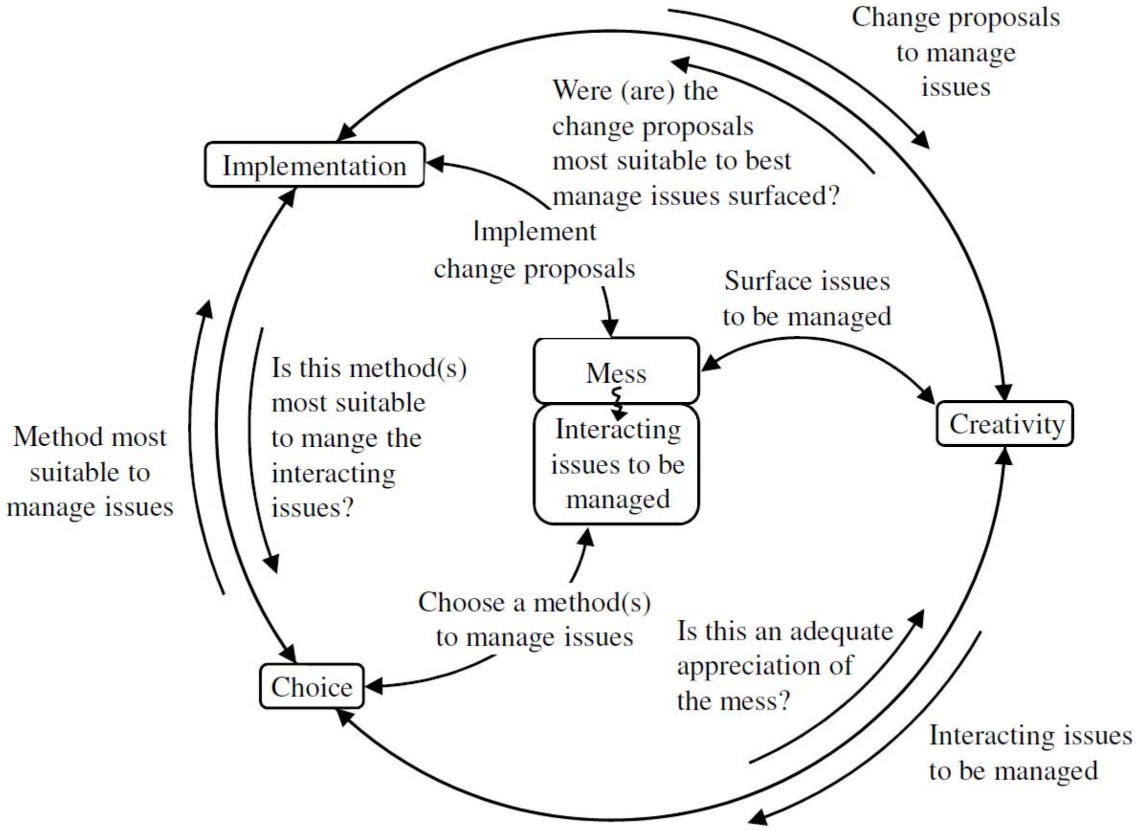


Figure II-9. The process of TSI in clockwise and counterclockwise modes [From Flood, 1995].

In contrast, the counterclockwise mode is a process of critical reflection in which the task of each phase is to question the outcome of the previous phase. For example, when the choice phase receives details of a set of interacting issues to be managed from the creativity phase, the critical reflection position asks, “Is this an adequate appreciation of the organization?” Consequently, each phase passes its outcome to the next phase in a clockwise direction and receives critical reflections about that outcome from the next phase in a counterclockwise direction. While this might suggest a sequential process, Flood (1995) stresses that no phase exists independently, but rather, all three phases occur simultaneously, albeit one phase may be in sharper focus at times compared to the

other two phases. Additionally, there exists a recursive structure with the TSI process whereby each of the three phases operates within all the other phases as a subphase.

The aim of the creativity phase is to identify issues to be managed by focusing on decontextualizing, contextualizing, and synthesizing the two. Decontextualization, which emphasizes divergent thinking that looks at the organization from many angles and viewpoints, provides the creative input necessary to identify a wide range of issues to be managed. Contextualizing then helps to converge on issues that should be managed and is guided by the use of systems metaphors as organizing structures for thinking creatively about the problem situation. These metaphors are machine, organic, neuro-cybernetic, socio-cultural, and socio-political. They respectively conceive organizations to be mechanistic, organic, organic but intelligent, as if they were a culture, or as if they were a political system (Table II-5). The first three metaphors focus on the technical activities of an organization, while the last two metaphors focus on the human activities. Implementation of the choice of issues then follows from the synthesis of decontextualization and contextualization, and selected issues to be managed are passed on to the choice phase (Flood & Jackson, 1991; Flood, 1995).

Table II-5. The main attributes of five metaphors used in the creativity phase of TSI [From Flood, 1995].

Metaphor	Main attributes
Machine	Standardized parts Routine operations Repetitive operations Activities predetermined Goals and objectives predetermined Efficiency Rational approach Internal control Closed system
Organic	Needs to be satisfied Survival Open system Adaptation Organization Feedback

Metaphor	Main attributes
	Self-regulation Passive control
Neuro-cybernetic	As organic, but also includes: • Active learning and control • Information prime • Law of requisite variety • Viable system • Learning to learn • Getting the whole into the parts
Socio-cultural	Collaboration Shared characteristics: • Language • History • etc. Shared reality • Values • Beliefs • Norms Social practices
Socio-political	Coercive conflict Domination Whose interests are served? Power central issue People are politically motivated Power as a consequence of structure Disintegration

Flood and Jackson (1991) proposed in their original description of TSI that the choice of systems methodology should be informed by the System of Systems Methodologies (SOSM). The SOSM attempts to logically group systems methodologies based on the underlying assumptions they make about problem contexts in terms of two dimensions: systems and participants. The systems dimension refers to the relative complexity of the system(s) that make up the problem situation, which are classified on a continuum of system types ranging from “simple” to “complex” (Table II-6).

Table II-6. Characteristics of simple and complex systems [After Flood & Jackson, 1991].

Simple	Complex
• Small number of elements • Few interactions between the elements • Attributes of the elements predetermined • Interaction between elements is highly organized • Well-defined laws govern behavior • The system does not evolve over time • Sub-systems do not pursue their own goals • The system is unaffected by behavioral influences • The system is largely closed to the environment	• Large number of elements • Many interactions between the elements • Attributes of the elements are not predetermined • Interactions between elements is loosely organized • They are probabilistic in their behavior • The system evolves over time • Sub-systems are purposeful and generate their own goals • The system is subject to behavioral influences • The system is largely open to the environment

The participant dimension refers to the relationship of agreement or disagreement between the individuals or parties who stand to gain or lose from a systems intervention. Participant relationships are classified as unitary, pluralist, or coercive based on the political characteristics of the situation in terms of the issues of interest, conflict, and power (Table II-7).

Table II-7. Characteristics of unitary, pluralist, and coercive relationships between participants [After Flood & Jackson, 1991].

Unitary	Pluralist	Coercive
• Share common values • Values and beliefs are highly compatible • Largely agree upon ends and means • All participate in decision making • Act in accordance with agreed objectives	• Have a basic compatibility of interest • Values and beliefs diverge to some extent • Do not necessarily agree upon ends and means, but compromise is possible • All participate in decision making • Act in accordance with agreed objectives	• Do not share common interests • Values and beliefs are likely to conflict • Do not agree upon ends and means and "genuine" compromise is not possible • Some coerce others to accept decisions • No agreement over objectives is possible given present systemic arrangements

Combining the dimensions of systems and participants yields a 6-celled matrix with problem contexts falling into the following ideal-type categories: simple-unitary, complex-unitary, simple-pluralist, complex-pluralist, simple-coercive, and complex-coercive. Each of these problem contexts differs in a meaningful way from the others, implying the need for at least six types of problem solving methodologies or systems approaches. Figure II-10 groups system methodologies based upon the assumptions they make about the nature of the problem context in terms of the systems from which problems or issues emerge and the participants. The systems methodologies included in the SOSM are not constrained to those shown in Figure II-10, and any well-formulated methodology may be included in TSI driven interventions (Flood, 1995).

	Unitary	Pluralist	Coercive
Simple	• Operations research • Systems analysis • Systems engineering • System dynamics	• Social systems design • Strategic assumption surfacing and testing	• Critical systems heuristics
Complex	• Viable system diagnosis • General system theory • Socio-technical systems thinking • Contingency thinking	• Interactive planning • Soft systems methodology	? ? ?

Figure II-10. Grouping of systems methodologies based upon the assumptions they make about problem contexts [From Flood & Jackson, 1991].

So where does HSI fit into the SOSM? Having asserted earlier in Section II-A that HSI is based on sociotechnical systems theory, Figure II-10 would imply that HSI is most applicable in complex-unitary problem contexts where there are shared values and agreement on ends and means. From a TSI perspective, in problems where this is not the case, then another (soft) systems approach is required to first make the problem more tractable for HSI. Likewise, in the case where the problem situation is complex, HSI can be applied to make complex-unitary problems simpler, and hence more tractable for hard systems approaches such as systems analysis and systems engineering. In this respect, HSI might be considered an enabler of these hard systems methods.

J. UNDERSTANDING HSI THROUGH TOTAL SYSTEMS INTERVENTION

1. Introduction

In Chapter I, we discussed the difficulty shared by many HSI practitioners, program managers, and engineers in answering the question, "What is HSI, and how should it work?" What now follows is a short case study that illustrates the use of TSI to address this very question. The focus is not on the main features of the method *per se*, but on the outcome in terms of a prototype systems model for "doing" HSI. Note that

this model is not offered as a universal answer to the original question. Rather, it summarizes the understanding gained from the application of soft systems approaches to a specific organization's problem of addressing human performance in systems. The application of soft systems approaches provides the necessary understanding of meanings vis-à-vis HSI such that a sensible attempt can be made in subsequent chapters to explore applying hard systems approaches to specific aspects of the problem. Thus, recalling Figure II-8, we are in effect choosing one of Pidd's (2004) suggested relationships between hard and soft systems approaches—namely, the right-hand part of the figure (reproduced below in Figure II-11)

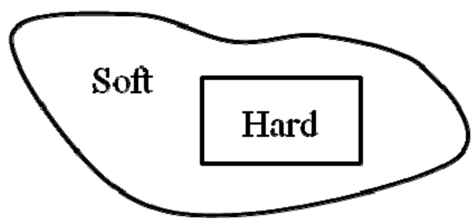


Figure II-11. Presumed relationship between hard and soft systems approaches [After Pidd, 2004].

2. Background

The armed services of the United States, though at core "a military force," are engaged in many research and development programs that entail advancing science and technology for the purpose of developing or improving systems crucial to military superiority. In transitioning "ideas to weapons" (Holley, 1953), the historical record suggests the need for the armed services to "develop effective systems for integrating the advances of science with the military machine" (Holley, 2004, p. 74). The armed services have responded by organizing large military laboratories and technical workforces that are responsible for a diverse portfolio of science and technology ranging from basic research to advanced technology development. The latter includes targeted research to shape the future battlespace, integrated technology options to satisfy identified requirements, and rapid solutions to meet urgent operational needs.

Starting in 1997, the United States Air Force has organized for research around the Air Force Research Laboratory, which consists of several technology directorates such as space vehicles, information, air vehicles, propulsion, directed energy, sensors, munitions, and human effectiveness. The latter directorate is composed of a diverse group of scientists and engineers who study developing technologies specific to the human element of warfighting capability. Their portfolio of science and technology projects predominately involve human factors engineering and training, and to a lesser extent, personnel selection and survivability. However, as a result of a major reorganization in 2005, the lab merged to form the 711th Human Performance (HP) Wing by combining its human effectiveness directorate with several external organizational entities responsible for aerospace medicine education, training, and consultation; occupational and environmental health; and HSI. This reorganization marked the creation of the first human-centric warfare wing to consolidate research, consultation, and education within a single organizational entity. The HP Wing's stated primary mission areas are aerospace medicine, science and technology, and HSI (reference http://www.wpafb.af.mil/afrl/711HPW/ accessed September 10, 2009).

Nearly one year after the organizational merger, the new HP Wing Director sent a message to his key managers. This message aimed to set the tone for the future growth and maturation of the HP Wing, and it did so in terms of developing "long term integrated Human Performance solutions to UAS [Unmanned Aircraft System] challenges including operator screening, selection, training, effectiveness, and other human systems integration considerations." In the spirit of the organizational culture it was trying to create, the message did not spell out exactly how the desired ends were to be achieved. Rather, the Human Performance Integration Directorate (HPID), the HP Wing's HSI function, was charged with developing a new approach to ensure that human performance challenges were considered in a more holistic manner.

Subsequent discussions between the HP Wing Director and HPID managers over a short period of time filled in the background to the message. Following multiple consolidations and downsizing of the Air Force laboratories in the 1990s, there was a shift in strategic direction away from improving human performance in complex system

operation by broadly addressing issues in selection, training, and equipment design to instead becoming a supplier of "technical solutions" to problems involving human factors engineering and training. This most recent major reorganization had resulted in the appointment of new top management, increased staff numbers, expanded expertise, and new missions. These changes imposed considerable demands upon both the HP Wing's management and workforce, and the HP Wing Director felt that the change process was far from complete.

In the volatile environment of UAS (Singer, 2009), the HP Wing had now acquired and/or was developing numerous science and technology projects supporting a range of stakeholders in the acquisition, training, and operational UAS communities. Yet, at the individual project level, the HP Wing's workforce remained oriented around particular domain competencies and pre-merger organizational identities. The HP Wing Director questioned how all their efforts fit together as part of a coherent strategy to address the broader range of UAS human performance considerations. He sought some unifying theme for holistically looking at the HP Wing's contributions that could persist as further changes occurred and as different organizational structures evolved. This search led to the idea of managing human performance-related science and technology projects using HSI as an embedded business practice. However, this new conception necessarily raised the question, what is HSI and how could HPID be the catalyst for organizational change? What the HPID managers needed was a broad approach to examine this problem situation in a way that could lead to decisions on action at the level of both "what" and "how."

3. Consulting Using Soft Systems Methodology Through TSI

a. TSI: Task, Tool, and Outcome

A consulting project was undertaken in June 2009, at the request of the HPID Director. It was agreed to hold a two-day participatory workshop at the HPID office in San Antonio, Texas to explore HPID's role in diminishing the HP Wing's difficulties addressing human performance in UAS. Managers from the principle HPID divisions as well as those in the organization from the UAS integrated product team were

convened for the workshop. The task was to learn how the HPID team perceived and understood their current problem situation. Given prior discussions with the HPID Director, it was determined that the problem context was best characterized as complex-pluralistic. Thus, the appropriate tool was soft systems methodology employed through TSI. The expected outcome was a set of recommendations for new ways of organizationally addressing HSI in the HP Wing.

b. Creativity

Once the HPID team assembled for the first day of the 2-day workshop, formal proceedings began with approximately two hours of unstructured but intense discussion of the HP Wing and its problems managing UAS-related projects. By the end of the second hour, the energy of the discussion had begun to subside as now-familiar complaints were being reexpressed. At this point, the HPID Director presented his view that the workshop should not simply focus on managing the current problem situation but should rethink HPID's role within the HP Wing and refine it. In effect, he was indicating to his managers that he was looking for a fundamental study not constrained to UAS-related matters.

The HPID team responded by first formally recording a "finding out" phase even though most managers felt they were well steeped in the issues. They readily perceived their overall situation as complex, with no simple, unitary definition of "the problem." Encouraged to name "problem themes," the team started with several key phrases from the HP Wing Director's message:

- "[Those] working on pieces of the UAS effort aren't talking to each other and coordinating efforts when they should be"
- "Question is [sic] do all the separate actions fit together as part of a coherent strategy to address the broad range of human performance considerations"
- "[Look] at our contributions holistically"

The HP Wing Director's expressions of concern had something in common: they implied a need for a planning and organizing function (i.e., a neuro-cybernetic finding). Two other dominant concerns then subsequently emerged. First, while the HPID team was in

agreement over their currently stated mission of "advocating, facilitating and supporting the application of HSI principles to optimize operational capabilities," there was an uneasy acknowledgement that they lacked a holistic appreciation of the necessary activities within HPID to achieve this purpose (i.e., an organic finding). Second, the HPID team noted that the other HP Wing organizations were essentially nets of semi-autonomous groups, each consisting of relatively independent professionals exercising their professional judgment in addressing what they perceived as unmet human performance needs. Many of these groups corresponded to clusters of experts in the human factors engineering (HFE); personnel (P); training (T); and environment, safety and occupational health (ESOH) domains of HSI. It was also evident to the team that many, if not most, HP Wing personnel worked *in* HSI (via one of the domains) but not *for* HSI, which is to say there was no unitary power structure for managing a HSI process within the HP Wing (i.e., a neuro-cybernetic and socio-political finding).

c. Choice

Given that the problem context was considered complex-pluralist, we chose to introduce soft systems methodology (SSM) to the HPID team. SSM grew out of the work by Checkland and colleagues at the University of Lancaster to apply systems ideas to tackle "messy" management problems (Checkland, 1981; Checkland & Scholes, 1990; Checkland, 2000; Checkland & Poulter, 2006). The basic premise underlying SSM is the argument that individuals or groups necessarily attribute meaning to their perceptions of the world. These meanings constitute interpretations of the world (i.e., "worldviews"), the latter derived from previously gained experience-based knowledge of the world. Such interpretations inform perceptions and intentions, which can be translated into purposeful action. Once taken, purposeful action changes the experienced world and the process repeats. This argument places knowledge acquisition in a cycle that embodies the possibility of learning, in which case purposeful action can be aimed at intended improvements. This process then lends to the idea of formally operating the learning cycle so that purposeful action is taken in specific situations to bring about what are deemed improvements by those carrying out the process.

SSM provides an organized way of approaching problematic situations based on this process of inquiry and social learning. It has evolved to take the form of the four-activities model or the inquiring/learning cycle illustrated in Figure II-12. Checkland (2000) defines the four activities as follows:

1) Finding out about a problem situation, including culturally and politically;
2) Formulating some relevant purposeful activity models, each made to encapsulate a declared worldview and comprised of a cluster of linked activities that together make up a purposeful whole;
3) Debating the situation, using the models, seeking from the debate both
   a) changes which would improve the situation and that are regarded as both desirable and feasible, and
   b) the accommodations between conflicting interests that will enable action-to-improve to be taken;
4) Taking action in the situation to bring about improvement.

In practice, it has been observed that the latter activities tend to take on one of two foci. The first is the original one in which SSM is an action-oriented approach, seeking the enabling accommodations necessary for any action-to-improve to be taken. The second more recent focus is on SSM as a sense-making approach in which activity models are put to use to improve understanding of complex situations.

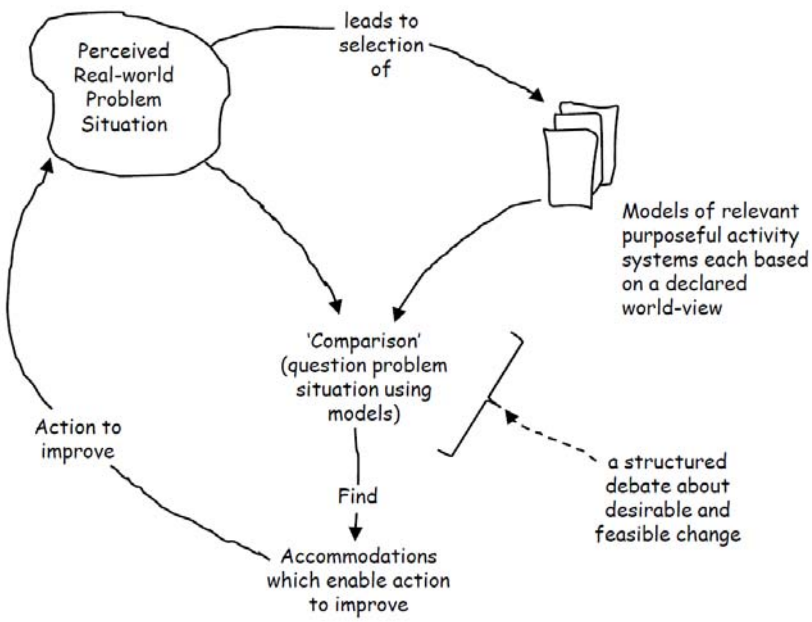


Figure II-12. The inquiring/learning cycle of soft systems methodology [After Checkland, 2000].

d. Implementation

Conceptual models are used in SSM as intellectual devices for ensuring an organized process of inquiry and learning. Model building begins with the formulation of the names of relevant systems, which must be crafted so as to make it possible to assemble a logic-based model of the systems named. These names become "root definitions" since they express the core or essence of the perception to be modeled. All root definitions express a purposeful activity in terms of a transformation process, T, in which some entity, the "input," is changed into some new form of the same entity, the "output." Well-formulated root definitions are prepared by consciously considering the elements captured in the mnemonic CATWOE, which stems from the initial letters of the following six terms:

- *Customers*: The victims or beneficiaries of T.
- *Actors*: Those who would do T.
- *Transformation process*: The conversion of input to output.
- *Weltanschauung*: The worldview that makes this T meaningful in context.

- *Owner(s)*: Those who could stop T.
- *Environmental constraints*: Elements outside the system that it takes as given.

The SSM modeling language is based upon verbs (i.e., activities), and the modeling process consists of assembling and structuring the minimum necessary number of activities to carry out a single transformation process in light of the definitions of the CATWOE elements. These activities are derived by means of a straightforward logic-based stream of analysis and the heuristic guideline is to aim for 7 ± 2 individual activities. Once enumerated, individual activities are linked together based on whether or not they are "contingent upon" or "logically dependent upon" another activity.

Accordingly, the first model produced by the HPID team represented a notional system that directly addressed the HP Wing Director's stated concerns. The root definition of this system is shown in Figure II-13; Figure II-14 shows the root definition and its CATWOE in conventional form. It was instantly appreciated that this EROS (i.e., Environment-Relation-Operation-Support) model was at too high a level of abstraction for describing relevant systems. However, setting the real world expressions of concern against this simple concept proved useful in provoking further discussion and insights. These insights provided the basis for producing the rich picture (Figure II-15) that served as the genesis for a more detailed model of a general (primary issue) system to satisfy human performance needs.

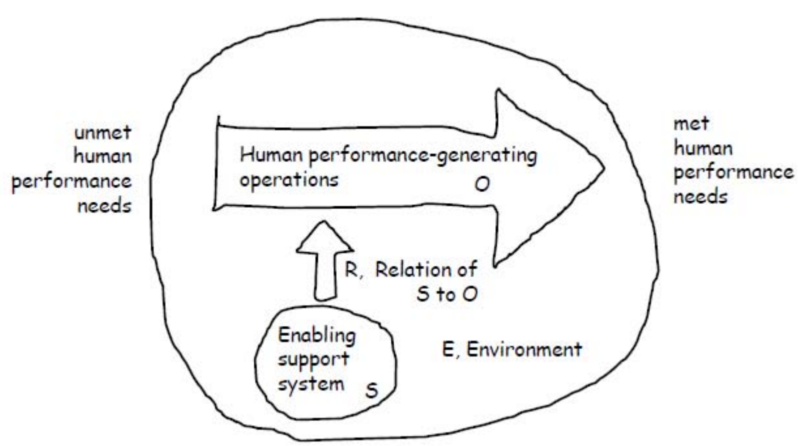


Figure II-13. A representation of HPID's role in the HP Wing (as one of the enabling support systems, S).

Root Definition

> A Wing-owned system staffed by HPID personnel to coordinate UAS human performance-focused research programs by integrating project planning, execution, & data synthesis to holistically address UAS human performance needs, both internally and with respect to external customers

C: Wing project staffs

A: HPID personnel – 'impartial' integrators

T: Coordination of projects

W: Coordination & synergies obtained by pulling together elements, findings & relationships from HFE/P/T domains in terms of HP trade-off surfaces (HSI process model)

O: Wing director

E: Strategic R&D plans, funding streams, external cust. demands

Figure II-14. Root definition and CATWOE based on Figure II-13.

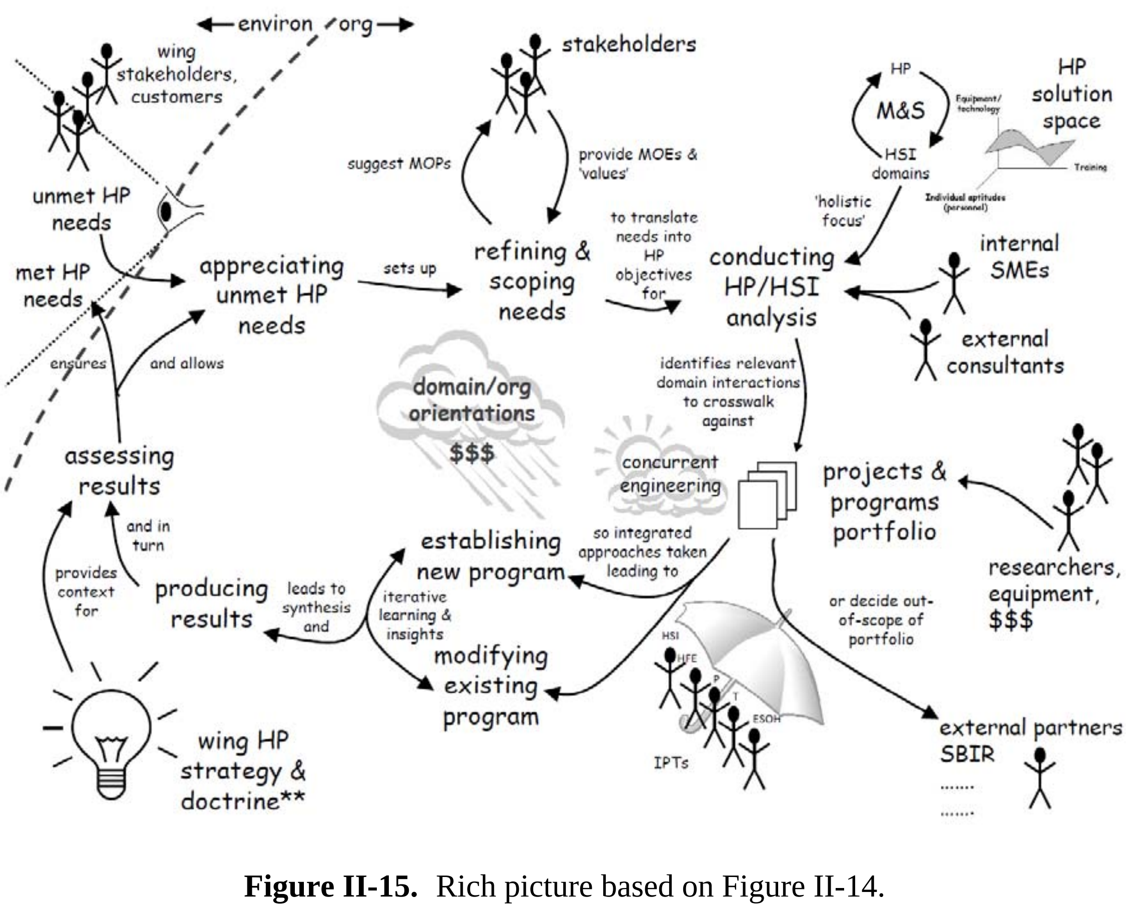


Figure II-15. Rich picture based on Figure II-14.

In the course of producing a rich picture, the HPID team discovered they were also building a structural picture of a larger problem situation, which led to yet another round of "finding out." Aided, no doubt, by the extremely simple nature of the EROS model of Figure II-13, the team changed their perceptions to consider the broader context of the recently formed HP Wing as a unique Air Force asset for planning and developing human performance *capabilities*. This shift led to consideration of the Defense Department's capabilities-based planning process, often described in terms of the DOTMLPF mnemonic {doctrine (D), organization (O), training (T), materiel (M), leadership (L), personnel (P) and facilities (F)}, as a relevant wider system, and hence another *Weltanschauung* that should be considered by the team. If the DOTMLPF paradigm defines the solution space for developing new national security capabilities and human performance is considered a form for providing such capabilities, how then does the EROS model relate to this higher-level system? In answering this question, the team produced Figure II-16, which builds on prior work by the team drafting conceptual frameworks for a human performance doctrine (Tvaryanas, Brown, & Miller, 2009).

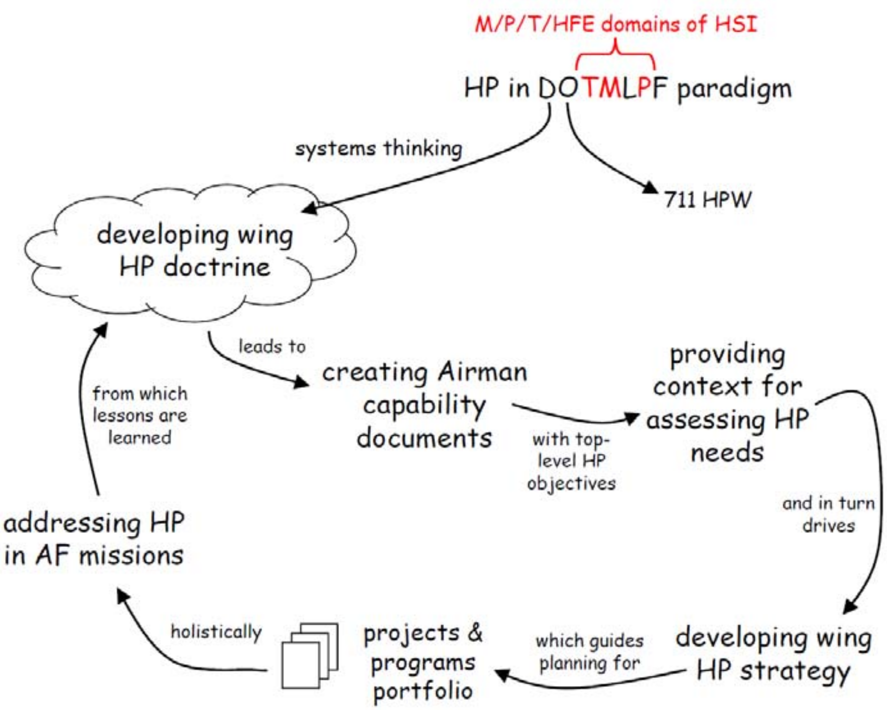


Figure II-16. Concept for systematically developing human performance-related capabilities.

The ensuing discussions provoked by Figures II-15 and II-16 enabled ideas for a relatively circumscribed number of activities to be agreed upon by the HPID team. These activities, depicted in the conceptual model shown in Figure II-17, described a purposeful activity system for meeting human performance needs. In a period of reflection, the team noted that the finding out phase had become very broad and was no longer limited to the original UAS-related expressions of concern. However, this issue was not in itself considered problematic as the team now appreciated that the UAS problem situation was actually just a specific manifestation of a larger HP Wing problem situation. Nevertheless, the team exercised the model using several UAS-related projects to reassure themselves that it was logically consistent.

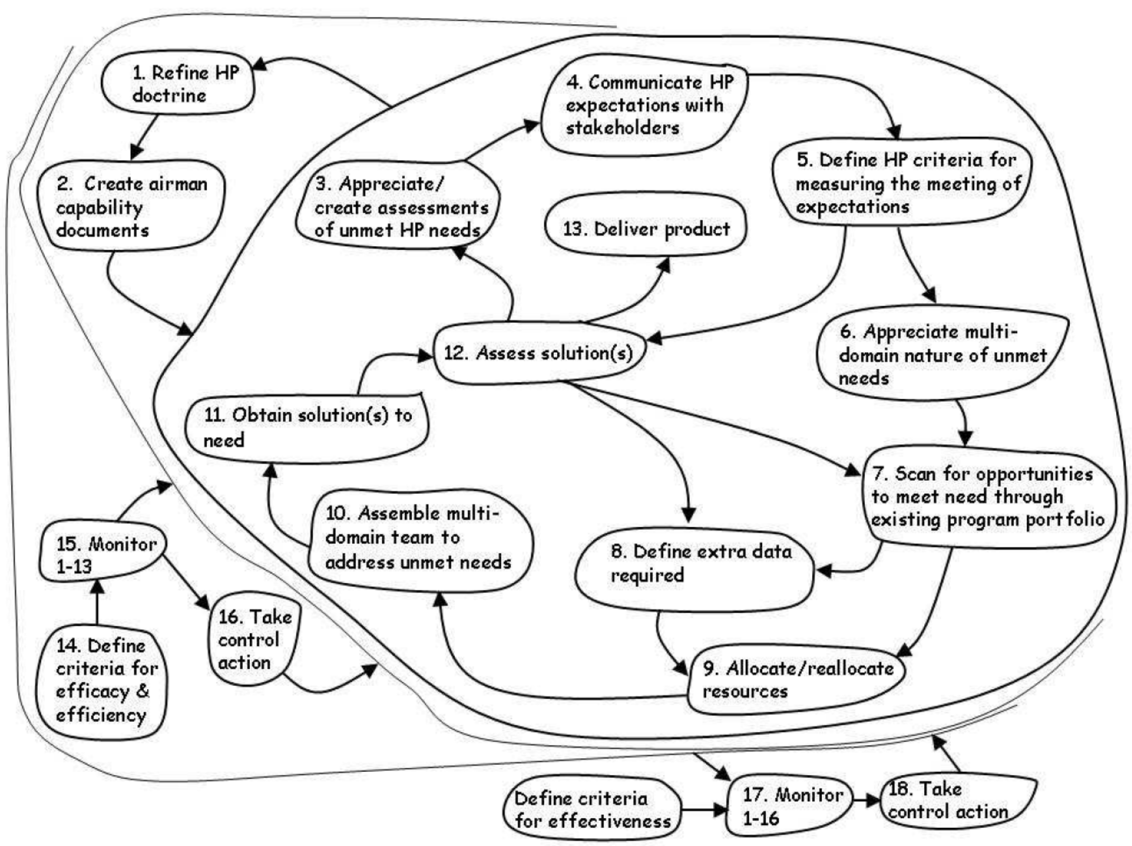


Figure II-17. The conceptual model for the root definition in Figure II-14.

The conceptual model in Figure II-17 was the first significant model produced by the HPID team, and it would prove to be one that raised much concern. As described earlier, the HPID Director desired to gain insight through this systems study into the role of HSI within a human performance-generating operation. However, at the start of the workshop, the HPID team was frustrated by its lack of success in articulating HSI as a transformational process in the sense of SSM in which "some entity" is converted into "that entity in a transformed state." The team's long held model of HSI (Figure II-18), which was used often in the past to good effect, envisioned the HSI domains as inputs and human performance as an output—not transformed HSI domains as SSM would have it.

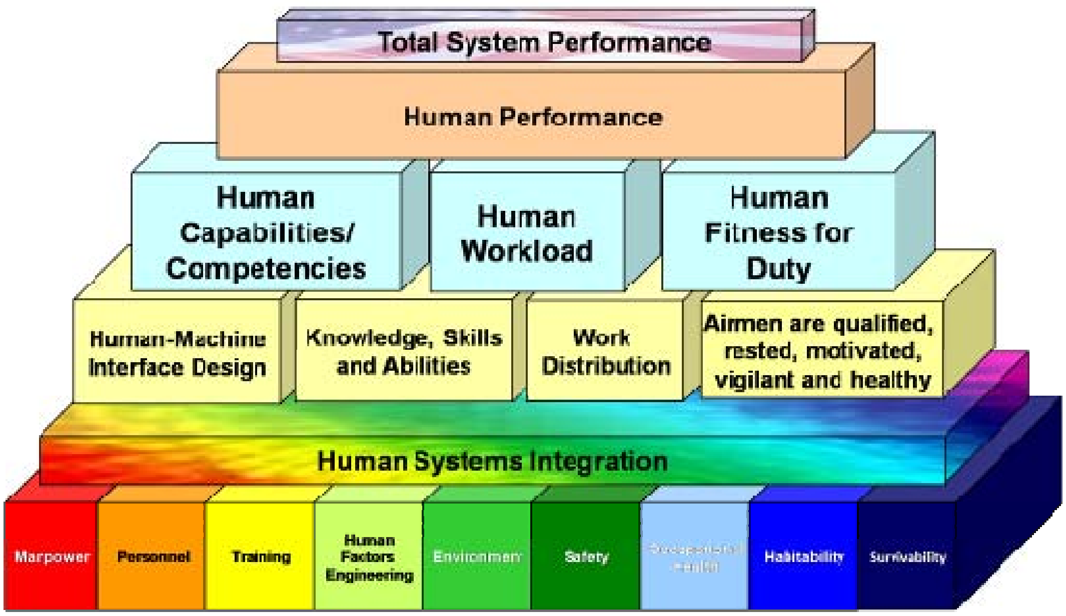


Figure II-18. Old Human Performance Integration Directorate HSI model.

Stymied in developing a root definition for HSI as a transformation process, the team had tabled the debate in favor of letting a description of the HSI process emerge from the analysis and model building directed at the HP Wing Director's expressions of concern. Now, at the end of that phase of the study, the team was dismayed to find that HSI was not a specifically named activity in the conceptual model

of Figure II-17. Slowly, the team came to appreciate that HSI, as they now understood it in relation to the conceptual model, addressed activities primarily focused on the planning and organizing, but not necessarily actually delivering, human performance solutions.

Even with this newfound appreciation, the conceptual model of Figure II-17 was at too high a level for gaining useful insights into HSI as a set of purposeful activities. The first thought was to expand each of the activities of the first conceptual model into an activity model at the next level of resolution. However, it was quickly concluded that this expansion would lead to too much detail and far too many activities with the potential for further obscuring the problem situation rather than providing clarification. Taking a more pragmatic approach, the HPID team decided to rank the 13 core activities in Figure II-17 to determine by consensus the activity they perceived as most representative of HSI. Activity 6 ("Appreciate multi-domain nature of unmet needs"), in itself a subsystem, was chosen as the framework for a second cycle of analysis and modeling.

The HPID team decided to approach Activity 6 using the original UAS-related expressions of concern, but agreed to work towards a broader context for the resulting conceptual model. Within the framework of Activity 6, HSI activities were conceptualized as transforming "human performance criteria" to "human performance criteria as multi-domain solution sets"—a correctly formulated transformation in the sense of SSM. Figure II-19 shows the concept and root definition (and its CATWOE) in conventional form for Activity 6.

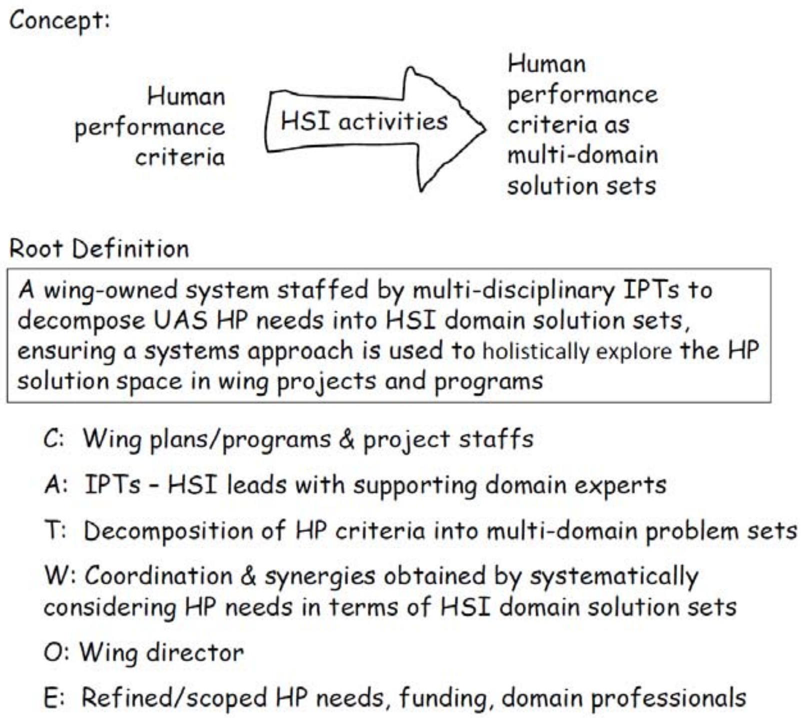


Figure II-19. Concept and root definition for Activity 6.

With the prior experience with the EROS model still fresh in memory, the HPID team set about placing the real-world UAS-related expressions of concern against the very simple transformation depicted in Figure II-19. The discussion this provoked enabled an intense "finding out" phase, which led to the production of a basic structural picture of the problem situation (Figure II-20) as well as a rather detailed model of a system to generate human performance solutions (Figure II-21).

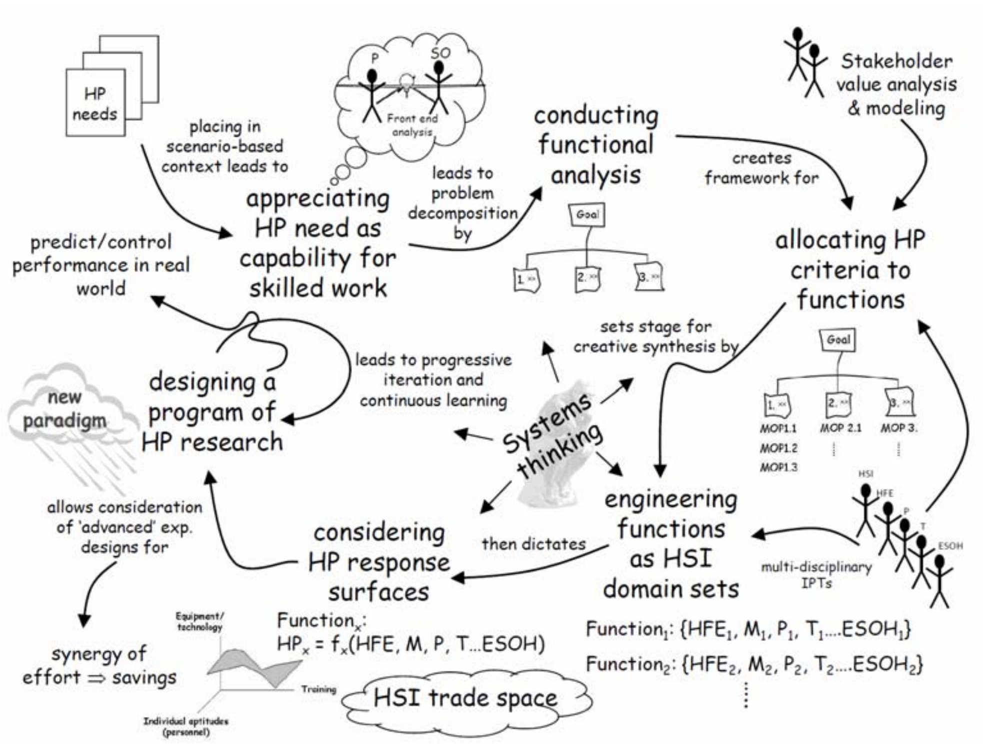


Figure II-20. A rich picture of a system concept for generating human performance solutions.

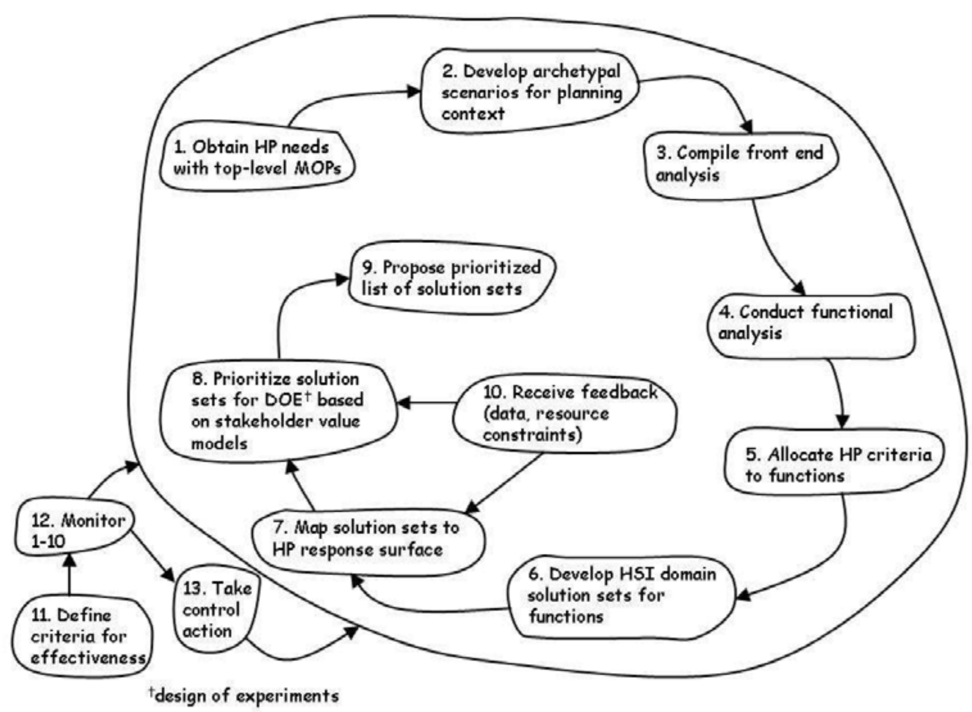


Figure II-21. Activity 6 of Figure II-17 expanded.

During their iterative elaboration, the system concept and conceptual model quickly shed explicit reference to the original UAS-related expressions of concern (with the sole exception of the UAS depicted in Figure II-20) in favor of a more generalizable system. Again, this was not considered problematic, particularly since this was the team's conscious intent. Nevertheless, it was judged prudent to exercise the model using several UAS-related projects to ensure that it was logically consistent.

In carrying out the second methodological cycle, there was an expected shift from the level of "what activities to do" to that of "how activities should be carried out." This shift reflected the transition in the HPID team's focus of inquiry from the original system level to the sub-system level. The result was a view of HSI that was significantly different from that previously held by the HPID team. This change occurred largely because the method used in this study drove the team back to the *raison d'être* of the HP Wing—as a human performance-generating operation—which subsequently led to radically different ideas about HPID's role and constituent activities relative to the larger organization. Initially, the HPID team had been unable to articulate HSI as a purposeful set of activities because they first needed to work at the higher levels of "why" and "what," rather than at the structural level, which is about "how." This point was illustrated poignantly in the systems concept shown emblematically in Figure II-20, which begins with "appreciating HP need as capability for skilled work"—a direct allusion to the capability paradigm embodied by the wider Defense Department system that contains the HP Wing. This insight was unavailable to the team at the start of the workshop.

Indicative of the HPID team's changing perceptions was their feeling that a more detailed level of analysis was needed for the conceptual model in Figure II-21 and some parts of the model in Figure II-17. However, it was also becoming clear that the team was coming to the end of the work that they themselves could do in the workshop. The detailed work accomplished up to this point enabled the team to build a preliminary account of a refined HPID and to examine logically and in detail the organizational changes required by the HP Wing to address the Director's expressions of concern. Reference to a wider audience was now necessary.

e. Conclusions From the Workshop Example

By the close of the workshop, the experience of the HPID team with TSI and SSM enabled them to sharpen their ideas about problematical issues and to do so holistically. It also introduced them to the idea of making models of purposeful activity systems and structuring debate. This learning was accomplished initially by means of the very simple EROS model, which led to more expansive thinking in terms of developing human performance capabilities for a wider system. In turn, this expanded thinking enabled insight into the need for a system to provide the environmental context for holistically assessing human performance needs and gauging whether or not they are met—a key concept for enabling hard systems approaches. This concept was initially, represented in Figure II-13 as the environment (E) supporting the human performance-generating operations (O). Later, the conceptual model in Figure II-17 showed development of this environmental context as a deliberate and controlled set of activities supporting the human performance-generating operations. This enabling role of a human performance doctrine was a particularly novel discovery. However, it was the conceptual model in Figure II-21 that really pointed the team towards a new vision of HPID as a proactive support function rather than as a reactive service function. The difference was between, on the one hand, *reacting to requests* for HSI consultation efficiently and effectively, and on the other, proactively *supporting the business* of human performance-generation through the planning and organizing activities of management.

As the HP Wing exists currently, domain-specific scientists develop and apply the fundamental theory that supports their domain, in essence determining what can be done in terms of generating domain solutions to meet human performance needs. These scientists are capable and eager for autonomy in providing solutions in accordance with their domain expertise and with a strong focus on their individual clients (and hence funding sources). Yet, based on the *Weltanschauung* of the HPID team, human performance solutions are actually solution sets, as illustrated in Figure II-20, which must be described in terms of each of the HSI domains. By virtue of this *Weltanschauung*, human performance should be considered as emerging from the amalgamation of domain-specific theories.

In the “more radical” view of HSI that formed from this study, the HPID team envisioned their workforce providing the theoretical synthesis across domain sets with the more pragmatic aim of satisfying multiple clients. To do this job, HPID personnel would need to understand each of the HSI domains, work effectively with their domain colleagues, and know when to bring interdisciplinary teams together to create solution sets, design studies, and interpret results. All together, the methods and concepts of SSM (through TSI) significantly improved the HPID team’s understanding of HSI. Without a doubt, this case study shows only one convoluted trajectory in understanding HSI, of which there probably could be any number. The details of this consultancy, however, offer a general lesson on applying systems methodologies within a TSI perspective to move our understanding of HSI from “messes” to interacting issues that can be managed.

K. BRINGING IT ALL TOGETHER

The discussion up to this point has attempted to tackle the issue of HSI as a philosophy or discipline. We traced the origins of HSI philosophy to the early human factors movement, which began in earnest during the Second World War and approached the problem of human performance in systems from the reductionist perspective of the scientific method. We then discussed the limitations of this approach in dealing with the complexity inherent in the problem, leading to the emergence of the systems-oriented disciplines of macroergonomics and HSI based on sociotechnical systems theory and the concept of joint optimization of personnel and technological subsystems. These disciplines are distinctly different from the other human factors-related disciplines in that there primary focus is a system, and hence they are artifacts of the transition in thinking about human performance in systems from the Machine Age to the Systems Age:

> HSI is not “post-modern” human factors; it is the evolution of human factors within the context of a larger systems movement that has occurred in response to the issue of irreducible complexity.

As part of a larger systems movement, HSI provides a systems account of the world and a systems approach to its problems. HSI thinking is, therefore, necessarily holistic and embraces the two pairs of ideas that are core to systems thinking in general:

emergence and hierarchy, and communication and control. In turn, the concepts of emergence and hierarchy lead to the consideration of human performance in systems as an emergent phenomenon that exists within a hierarchy of complexity. The concepts of communication and control drive a dynamic view of human performance from the perspective of complex adaptive systems. A major premise of the systems movement at large is that these ideas will enable us to tackle problems that the traditional scientific method has found difficult to resolve. However, much work remains in the human factors-related sciences to explore the consequences of this shift to holistic rather than reductionist thinking.

Since sociotechnical systems theory deals with optimization of personnel and technological subsystems within an organization, by implication, HSI "lives" relatively high in the hierarchy of complexity. This idea was explored in terms of several systems typologies, which made it evident that the task of joint optimization involves the integration of very different types of subsystems. We also considered the implication of dealing with problems that emerge relatively high in the hierarchy of complexity. Such problems, referred to as "messes" by Pidd (2003) or "wicked problems" by Rittel and Webber (1973), entail evolving sets of interlocking issues and constraints that can be managed, but not solved, and for which there are many stakeholders with divergent values who must be satisfied. Such problems often must be tackled using soft systems approaches to make them more tractable for hard systems approaches, the latter being the tools and methods of systems analysts and systems engineers. Flood and Jackson (1991) provide a meta-methodology (i.e., TSI) that deals with messes by first creatively identifying issues to be managed and then choosing and implementing the systems method(s) that are best equipped to tackle those issues most effectively. They provide a logical organization of systems methods, which suggests various soft systems approaches should first be employed to make problems more tractable for HSI in terms of clarity of objectives. In turn, HSI can then be used to make problems more tractable for hard systems approaches by reducing problem complexity.

L. AN HSI HYPOTHESIS

At the opening of this chapter, we looked at several logical definitions of HSI based on combinations of the definitions of the three constituent words: human, system(s), and integration. We later presented a case study in which SSM (through TSI) was used to help an organization understand HSI as a purposeful activity. This study was inspired, in no small part, by a similar study conducted by Checkland (1981) in which SSM was used to help clarify the theoretical concept of "terotechnology":

> In the 1970s some of the Government money spent to help improve industrial efficiency in the UK was channeled to what are known as the 'industrial technologies'. These were originally conceived as 'multi-disciplinary technologies', applicable to many different industries, whose neglect led to economic inefficiency. In the early 1970s, there were four of them: corrosion technology, tribology, materials handling technology, and 'terotechnology'. We were asked to help define the latter. In 1972, there was a newly constituted Committee of Terotechnology but no agreed definition of the concept. The Committee set up a Panel…to propose an argued definition, indicating exactly what was within the concept and what was not…The problem situation was an interesting one in that it was entirely arbitrary. It was not a case of defining and describing something which existed in the real world. Rather the task was to define a concept which in the opinion of the Department of Trade and Industry and some interested industrialists ought to be taken seriously by anyone concerned with the process of generating wealth by industrial activity (p. 202).

Their officially sanctioned definition of terotechnology follows (for those familiar with the Defense Department's policy guidance on HSI, it is interesting to note a number of shared themes between terotechnology and HSI!):

> Terotechnology is a combination of management, financial, engineering, and other practices applied to physical assets in pursuit of economic life-cycle costs. Its practice is concerned with the specification and design for reliability and maintainability of plant, machinery, equipment, buildings and structures, with their installation, commissioning, maintenance, modification and replacement, and with the feedback of information on design, performance and costs (p. 205).

In the case study of the HP Wing (which also used SSM), we might define HSI, at least as it was perceived by the study participants with regards to their processes, as follows:

> HSI is a system staffed by multi-disciplinary integrated product teams that decomposes human performance needs, identified by internal and external customers, into HSI domain solution sets for the purpose of strategically designing programs of research to systematically explore the entire human performance trade space.

This definition is obviously very different from the HSI definition offered by the National Research Council (2007):

> HSI [refers] to the design activities associated with ensuring that the human-system domains…are described in concert with all the other design activities associated with the systems engineering process, so that the resulting designs are truly responsive to the needs, capacities, and limitations of the ultimate users of the systems (p. 11).

We could continue citing or deriving definitions, but such a list would be quite long and exhibit a high degree of variability, solving little.

At the end of the HP Wing SSM case study, I stated that there are likely an innumerable set of definitions of HSI that could be elaborated based on one's *Weltanschauung*, so it is senseless to argue that I have a universal definition. I have also avoided, at least up to this point, enumerating a list of HSI domains, which like definitions, appear to vary between organizations. For example, the Canadian armed forces describe five HSI domains, the United Kingdom Ministry of Defense uses six domains, and the U.S. Defense Department has seven domains—although some military services within the Defense Department list more than seven. As was discussed previously in regards to the division of knowledge into scientific disciplines, the division of HSI into domains is necessarily man-made and arbitrary and largely a matter of organizational convenience. So again, I do not argue for an exhaustive and mutually exclusive set of HSI domains. Rather, I will simply declare my *Weltanschauung* as that of the Defense Department, and in so doing, I will accept their arbitrary set of seven HSI domains as listed in DoD Instruction 5000.02 (2008): human factors engineering, personnel, habitability, manpower, training, safety and occupational health, and survivability.

Now, instead of proceeding directly to a definition, I start with the concept of the prime directive. Hitchins (1992) asserts that the prime directive, which is the highest

level of abstract, objective statement of a system's purpose, is central to the idea of conceiving systems. He suggests four features that characterize a good prime directive: highest level of abstraction, ultimate purpose, sphere of endeavor, and solution transparency. Accordingly, I offer the following prime directive for an HSI system:

> To produce sustained system performance that is humanly, technologically, and economically feasible.

While terse, this prime directive succinctly expresses the *raison d'être*, the limits of action, and the sphere of activity of HSI. It does not over-specify; "produce" is vague yet entirely sufficient for purpose and there is no hint of solution in the prime directive's wording. According to the Defense Department *Weltanschauung*, HSI must ensure that the technological subsystem can accommodate the characteristics of the people in the personnel subsystem. It also should drive towards the most cost-effective solution in terms of life-cycle costs (DoD, 2008). These two notions are both Defense Department mandates for HSI as well as constraints on the potential solution space. Additionally, they are both called out specifically in the HSI prime directive, making our approach to HSI commensurable with official Defense Department policy.

Given the HSI prime directive, it is possible to proceed by semantic analysis to a supporting definition of HSI. Semantic analysis is a straightforward process in which each word in a statement, in this case the HSI prime directive, is examined and expanded to extract all meaning that it might contain, whether stated or implied. In so doing, and with an implicit reference to sociotechnical systems theory, we arrive at the following definition of HSI:

> HSI is a philosophy applied to personnel and technological subsystems within organizations in pursuit of their joint optimization in terms of maximally satisfying organizational objectives at minimum life cycle cost. Its practice is concerned with the specification and design for reliability, availability, and maintainability of both the personnel and technological subsystems over their envisioned life cycle.

This definition expands our understanding of the prime directive by providing increased discrimination with regards to our strategy for its achievement. The major new insights include:

1) HSI is a *philosophy* concerned with the joint optimization of the personnel and technological subsystems (i.e., sociotechnical systems theory) comprising some system-of-interest (SOI). These subsystems are being optimized with regard to some emergent property, which can only be observed from the level of the SOI's containing system (Figure II-22). The personnel subsystem is generally the province of human resources management and the technological subsystem is often the realm of some type of technical or engineering management. Consequently, the layer of organizational management within the containing system with cognizance over both the human resources and technical managers is the appropriate entity to address joint optimization, and hence HSI.

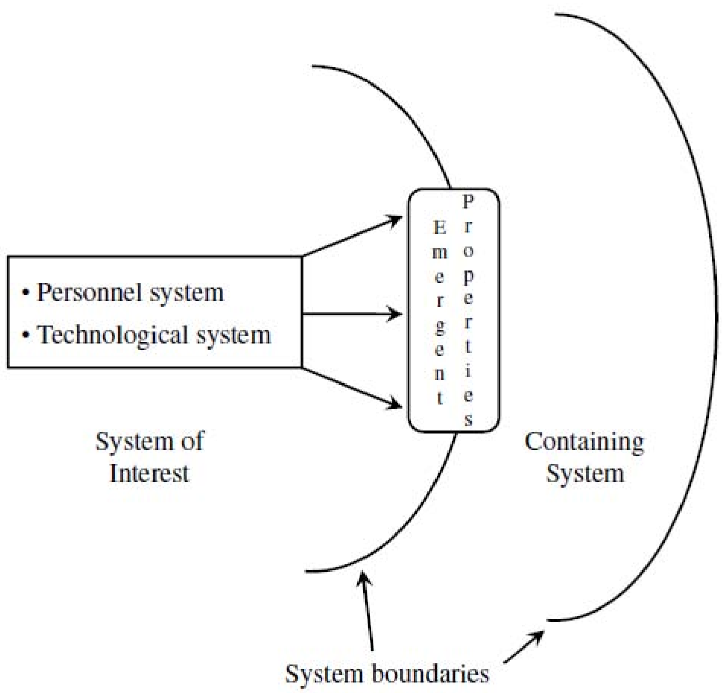


Figure II-22. Personnel and technological subsystems comprising system-of-interest (SOI) as viewed from that SOI's containing system.

2) HSI must continuously address the issue of the *sustained performance* of a SOI over its life. Organizations operate systems to perform functions required to achieve objectives that are believed to further organizational goals. HSI, through

joint optimization, is concerned with the properties and interactions of the personnel and technological subsystems such that the emergent properties of the SOI meet the objectives specified by those in the containing system. Usually, organizations desire that these emergent properties are maintained through time—they want the system to work today, tomorrow, and possibly the next year, decade, etc. Given the concept of joint causation, HSI must then be concerned with changes in the SOI's environment and corresponding adaptive changes to its subsystems (recall the earlier discussion on complex adaptive systems). Hence, joint optimization may be short lived, necessitating that the issue be continuously managed rather than definitively solved. Recall the prior discussion of wicked problems!

3) The focus on sustained performance naturally leads to a concern with designing for operational feasibility, meaning that the system will perform as intended in an effective and efficient manner (Blanchard & Fabrychy, 2006). In terms of the technological subsystem, this concern includes such design dependent parameters as reliability, availability, and maintainability. Note, however, that our HSI definition extends these same concepts to the personnel subsystem. It remains to be shown how this extension can be done, and we will begin to address this task in Chapter V starting with reliability. Nevertheless, in applying these concepts to both subsystems, it then becomes possible to examine their joint optimization within the SOI using a common set of constructs.

4) HSI can be viewed from both a *local or global perspective*. For example, consider Figure II-23, depicting a SOI and three sibling systems, each consisting of a personnel subsystem and a technological subsystem. The SOI and sibling systems, in turn, are components of a larger containing system (this is simply another example of a hierarchy of complexity). Assuming a local HSI perspective, we would seek the joint optimization of the personnel and technological subsystems within the SOI, thereby increasing the SOI's effectiveness and contributing positively to the containing system's objectives. This view is the traditional approach to HSI as it is applied in a large Defense

Department weapon system acquisition. Now, let us assume that the SOI and its sibling systems must share a common personnel resource pool. It is possible in optimizing the SOI to have unintended downstream effects on the personnel subsystems of the sibling systems. These downstream effects could result in decreased effectiveness of the sibling systems. In aggregate, optimizing the SOI may actually result in a net negative contribution towards achieving the containing system's objective! Such a scenario illustrates the need to also consider a global HSI perspective.

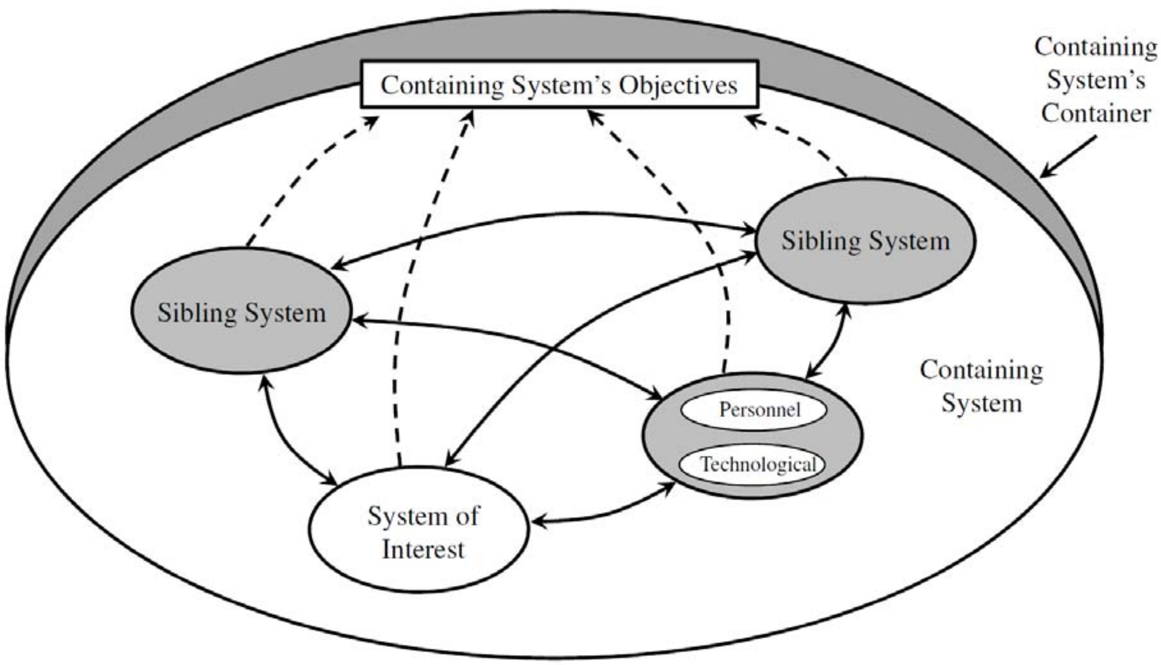


Figure II-23 A family of interacting systems, to include a system-of-interest and its sibling systems, all existing within the environment provided by the containing system.

These two perspectives, local and global, differ both in their level of focus within the hierarchy of systems and their metrics for assessing the worth of systems. The local perspective lends itself to optimizing the SOI in terms of effectiveness, which addresses the question, "Which of the proposed solutions is best?" In contrast, the global perspective addresses net contribution, which asks the question, "How do the emergent

properties of the SOI contribute to its containing system?" The global perspective, being based on net contribution, offers the following advantage:

> If all systems were evaluated correctly using Net Contribution, and only net positive solutions accepted, then—owing to the recursive nature of the technique, a hierarchy of net positive systems contained within next positive systems must develop. Thus, Net Contribution presents a high degree of implicit integrity in its effects on environment, its use of resource and its development of effective, enduring systems (Hitchins, 1992, p. 109).

However, these benefits come at the cost of greatly increased complexity in the process of evaluating options. Fortunately, our definition of HSI allows either perspective!

M. FINAL THOUGHTS

As the title states, this chapter provides a brief introduction into an admittedly rudimentary HSI philosophy, a topic that has yet to receive substantive treatment elsewhere. In suggesting the notion of a HSI philosophy, there is the implicit assumption that the subject of "HSI" aspires to the status of a serious discipline. It is probably fair at this point in time to characterize HSI, at best, as an emerging discipline. Accordingly, this chapter has attempted to bring together the views and concepts from a variety of systems thinkers and present a new set of ideas for perceiving, understanding, and analyzing problems from a unique HSI *Weltanschauung*. This attempt has a bold aim and it is difficult to prove or disprove any of the contentions presented, which is why, in the end, an HSI *hypothesis* is offered rather than a *theory*. It is open to criticism on that score, but hopefully in the process, it advances thinking on the nature of HSI as a discipline.

N. BIBLIOGRAPHY

Ackoff, R. L. (1981). *Creating the corporate future*. New York, NY: John Wiley and Sons.

Bertalanffy, L. von (1972). The history and status of general systems theory. *The Academy of Management Journal*, 15(4), 407−426.

Blanchard, B. S., & Fabrycky, W. J. (2006). *Systems engineering and analysis* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Booher, H. R. (1990). Introduction: the MANPRINT philosophy. In H. R. Booher (Ed.), *MANPRINT: an approach to systems integration* (chap. 1, pp. 1–16). New York, NY: Van Nostrand Reinhold.

Booher, H. R. (2003). Introduction: human systems integration. In H. R. Booher (Ed.), *Handbook of human systems integration* (chap. 1, pp. 1−30). New York, NY: John Wiley and Sons.

Boulding, K. E. (1956). General systems theory – the skeleton of science. *Management Science*, 2(3), 197−208.

Brooke, Donald (2009, December 7). Proliferation of robots fuels liability questions. *San Jose Mercury News*, p. C4.

Chapanis, A., Garner, W. R., Morgan, C. T., & Sanford, F. H. (1947). *Lectures on man and machines: an introduction to human engineering* (Report No. 166-I-19). Baltimore, MD: The John Hopkins University, Department of Psychology. (DTIC Accession No. AD0639366)

Checkland, P. (1981). *Systems thinking, systems practice.* Chichester, U.K.: John Wiley and Sons.

Checkland, P, & Scholes, J. (1990). *Soft system methodology in action*. Chichester, U.K.: John Wiley and Sons.

Checkland, P. (2000). Soft systems methodology: a thirty year retrospective. *Systems Research and Behavioral Science*, 17, S11−S58.

Checkland, P., & Poulter, J. (2006). Learning for action: a short account of soft systems methodology and its use for practitioners, teachers, and students. Chichester, U.K.: John Wiley and Sons.

Comte, A. (1865). *A general view of positivism*. (J. H. Bridges, Trans.). London, UK: Trübner and Co. (Original work published 1848).

Deal, S. (2007, January). *Definition of human systems integration*. HSI working group meeting conducted at International Council of Systems Engineering International Workshop 2007, Albuquerque, New Mexico.

Department of Defense (2008). *Department of Defense instruction 5000.02: operation of the defense acquisition system*. Washington, DC: Author.

Descartes, R. (1637). *Discourse on the method of rightly conducting the reason, and seeking truth in the sciences*. Retrieved May 23, 2009, from http://www.literature.org/authors/descartes-rene/reason-discourse/index.html

Flood, R. L. (1995). Total system intervention: a reconstitution. *Journal of the Operational Research Society*, 46, 174−191.

Flood, R. L., & Jackson, M. C. (1991). *Creative problem solving: total systems intervention*. Chichester, U.K.: John Wiley and Sons.

Hawking, S. W. (1988). *A brief history of time from the big bang to black holes*. New York, NY: Bantam Books.

Hendrick, H. W. (1995). Future directions in macroergonomics. *Ergonomics*, 38, 1617−1624.

Hendrick, H. W. (2002). An overview of macroergonomics. In H. W. Hendrick & B. M. Kleiner (Eds.), *Macroergonomics: theory, methods, and applications* (chap 1, pp. 1−23). Mahwah, NJ: Lawrence Erlbaum Associates.

Hitchins, D. K. (1992). *Putting systems to work.* Chichester, U.K.: John Wiley and Sons.

Holland, J. H. (1995). *Hidden order: how adaptation builds complexity*. Reading, MA: Perseus Books.

Holland, J. H. (1998). *Emergence: from chaos to order*. Reading, MA: Perseus Books.

Holley, Jr., I. B. (1953). *Ideas and weapons*. New York, NY: Yale University Press.

Holley, Jr., I .B. (2004). *Technology and military doctrine: essays on a challenging relationship*. Maxwell Air Force Base, AL: Air University Press.

Human. (2009). In *Merriam-Webster Online Medical Dictionary*. Retrieved December 19, 2009, from http://www.merriam-webster.com/medical/human

Integrate. (2009). In *Merriam-Webster Online Dictionary*. Retrieved December 19, 2009, from http://www.merriam-webster.com/dictionary/integrate

Jordan, N. (1968). *Themes in speculative psychology* (vol. 5). London, U.K.: Tavistock Publications

Kennedy, R. S., Jones, M. B., & Baltzley, D. R. (1989). *Regions of isoperformance: an interactive software package for trading off training, personnel, and equipment* (Report No. AAMRL-TR-89-003). Orlando, FL.: Essex Corporation.

Kleiner, B. M. (2008). Macroergonomics: work system analysis and design. *Human Factors*, 50(3), 461−467.

Lyons, M. (2004). Insights from complexity: organizational change and systems modelling. In M. Pidd (Ed.), *Systems modelling: theory and practice* (chap. 2, pp. 21−44). Chichester, U.K.: John Wiley and Sons.

Meister, D. (1999). *The history of human factors and ergonomics*. Mahwah, NJ: Lawrence Erlbaum Associates.

Muchinsky, P. M. (1993). *Psychology applied to work: an introduction to industrial and organizational psychology*. Belmont, CA: Brooks/Cole Publishing Company.

Mueller, M. W. (2008, April). Human systems integration – what's it all about? *Insight*, 2(11), 7−10.

Pew, R. W., & Mavor, A. S. (Eds.) (2007). *Human-system integration in the system development process: a new look*. Washington, DC: National Academies Press.

Pidd, M. (2003). *Tools for thinking*. Chichester, U.K.: John Wiley and Sons.

Pidd, M. (2004). Complementarity in systems modelling. In M. Pidd (Ed.), *Systems modelling: theory and practice* (chap. 1, pp. 1−20). Chichester, U.K.: John Wiley and Sons.

Popper, K. R. (1957). *The poverty of historicism*. London, U.K.: Routledge and Kegan Paul.

Rittel, H. W. & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, 4, 155−169.

Simon, C. (1976). *Analysis of human factors engineering experiments: characteristics, results, and applications* (Report No. AFOSR-TR-77-0333). Westlake Village, CA: Canyon Research Group. (DTIC Accession No. ADA038184)

Simon, H. A. (1982). *Models of bounded rationality*. Cambridge, MA: MIT Press.

Singer, P. W. (2009). *Wired for war, the robotics revolution and conflict in the 21st century*. New York, NY: The Penguin Press.

System. (2009). In *Merriam-Webster Online Dictionary*. Retrieved December 19, 2009, from http://www.merriam-webster.com/dictionary/system

Tvaryanas, A. P., Brown, L., & Miller, N. L. (2009). Managing the human weapon system: a vision for an Air Force human performance doctrine. *Air and Space Power Journal*, 23(2), 34−41.

Wertheimer, M. (1938). Gestalt theory (W. D. Ellis, Trans.). In W. D. Ellis, *A source book of Gestalt psychology*. New York, NY: Harcourt, Brace and Co. (Original work published 1924).

III. A BRIEF HISTORY OF THE EMERGENCE OF THE DEFENSE DEPARTMENT'S HUMAN SYSTEMS INTEGRATION PROGRAM

> Edward Luttwak's study on specialized light units pointed out that while the armies of America's allies tended to be "equipment constrained," the U.S. Army was more "manpower constrained" (Romjue, 1993, p. 27).

A. PHILOSOPHY VS. PROGRAM

In Chapter II, we considered human systems integration (HSI) as a philosophy that emerged as a result of the limitations of traditional science in dealing with the complexity of human performance in systems. The general intent was to provide some insight into how one might think broadly about HSI as a philosophy absent the baggage of its programmatic instantiation in the real world. Such a statement necessarily implies that there is a distinction between HSI as a "philosophy" versus a "program" (Booher, 1990, pp. 3–5). HSI philosophy can be applied to any purposeful human activity; hence, it is organizationally agnostic. In contrast, HSI programs apply to specific organizations and are tailored to their individual ways of doing business. For example, the Defense Department's HSI program applies only to the Defense Department. However, the Defense Department's HSI program, as the first large-scale programmatic instantiation of HSI philosophy, has become a relative benchmark for discussions of the topic of HSI at large. This assertion is supported by Deal's (2007) observation that many HSI definitions are traceable to Department of Defense (DoD) Instruction 5000.2. Thus, the historical analysis which follows provides an ancillary but indispensable explanation of the concept of HSI as it came to exist within the U.S. military.

B. SOME CONTEXT

1. American Post-World War II Political Culture

The idea of building military systems to optimize the collective performance of the soldier and their weapon is not new. Wu Ch'i (430 – 381 B.C.), a recognized expert

on warfare whose name is frequently associated with Sun Tzu, author of *The Art of War*, is reported to have declared to the Marquis (i.e., nobleman) Wen of Wei (Sun Tzu, 1963):

> At present, My Lord, during the four seasons you cause animals to be skinned and lacquer their hides and paint them vermilion and blue. You brilliantly decorate them with rhinoceros horn and ivory.
>
> If you wear these in the winter you are not warm, and in the summer, not cool. You make spears twenty-four feet long and short halberds of half this length. You cover the wheels and doors of chariots with leather; they are not pleasing to the eyes, and when used for hunting they are not light.
>
> I do not comprehend how you, My Lord, propose to use them.
>
> If these are made ready for offensive or defensive war and you do not seek men able to use such equipment it would be like chickens fighting a fox, or puppies which attack a tiger. Though they have fighting hearts, they will perish (pp. 151–152).

This quote demonstrates that concern for integrating soldier and weapon was by no means a unique phenomenon of the 20th century. However, the cognizance of a "man/machine interface crisis" by senior military leaders in the 1980s, coupled with a wider organizational sense of urgency to systematically address the issue, culminating in major organizational change in the form of a Defense Department HSI program could be reasonably characterized as a unique phenomenon within the annals of military history. If we wish to then study this phenomenon, which is now one of history, we must accept *a priori* that we will not be able to do so in an entirely objective manner. As mentioned in an earlier chapter, Popper (1957) asserts that the best we can hope to accomplish is to write a history that is consistent with a particular point of view. Thus, Popper would have us, if possible, clearly articulate the point of view we are choosing. Accordingly, our intent is to sketch an analysis that provides an explanation for the development of a Defense Department HSI program in terms of America's post-World War II political culture and the resulting linkage between U.S. foreign policy, military strategy, and high technology.

We begin by looking at the work of Paul Edwards (1988, 1996) who proposed a cultural and historical accounting for the U.S. military's fascination with computing. In what follows, we borrow heavily from Edwards' work, expanding his premise from

computers to more broadly considering technology (computers being ubiquitous in modern technology). Edwards (1988) asserts that it is not possible to analyze the military's technological choices without an understanding of the larger political context and vice versa:

> Thus the *worldview* of military institutions and their *technological choices* are bound up. In other words, the tasks assigned to the U.S. military by the political process determine the types and quantities of technology it develops and deploys. But, at the same time, the available technologies also affect which assignments it believes itself ready to accept [emphasis in original] (p. 247).

Hence, if technological choices led to soldier and weapon integration problems in the 1980s, we must set about seeking to understand the prevailing political situation at the time. Accordingly, we next consider several key elements of the post-World War II U.S. political culture that Edwards ascribes as shaping the worldview of the American military. These elements are 1) the apocalyptic struggle with the former Union of Soviet Socialist Republics, 2) the long history of antimilitarist sentiment in American politics, and 3) the rise of technology-based military power (Edwards, 1988, p. 245).

In the collective American psyche, World War II was a "good war" that was fought against nationalist aggressors and the antidemocratic fascist ideology, a fact that was only reinforced by postwar revelations of Nazi atrocities. Given postwar Soviet maneuvering in Eastern Europe and the openly expansionist Soviet ideology, the American sense of a Biblical struggle between good and evil did not simply fade away after the war. Instead, it underwent transference with Stalin being equated to Hitler and communism replacing fascism as a total enemy, thereby facilitating the transition into the Cold War (Figure III-1). This transference also carried with it the World War II sense of conflict on a global and total scale. The Truman Doctrine of containment and worldwide American military support for "free peoples who are resisting attempted subjugation by armed minorities or by outside pressures" (Harry Truman as quoted in Compston & Seidman, 2003, p. 194) codified the continuation of global conflict more or less on a permanent basis. However, the U.S., at the time having only recently emerged from the economic depression and political isolationism of the 1930s, had no immediately

available models for its new global role other than those of World War II itself. Consequently, the key events of World War II became basic icons in the organization of American Cold War foreign policy and military strategy, from Munich (the danger of appeasement) to Pearl Harbor (always be prepared for surprise attack) to Hiroshima (victory through technologies of overwhelming force). Thus, Edwards asserts that the Cold War was not a new conflict with communism, but rather, it was a continuation of the American experience with the apocalyptic struggle of World War II, only projected onto a different enemy (Edwards, 1988, pp. 247–248).[2]

---

[2] A nice illustration of this element of Edwards' thesis is provided in Robert McNamara's memoir, *In Retrospect: The Tragedy and Lessons of Vietnam* (1995). In his recounting of the deliberations leading up to the 1965 decision to escalate U.S. involvement in Vietnam, McNamara wrote:

> …I want to quote [Secretary of State David Dean Rusk's] exact words, because his view—that if we lost South Vietnam, we increased the risk of World War III—influenced others of us to varying degrees as well. [Dean] wrote:
>
> > The integrity of the U.S. commitment is the principal pillar of peace throughout the world. If that commitment becomes unreliable, the communist world would draw conclusions that would lead to our ruin and *almost certainly to a catastrophic war*. So long as the South Vietnamese are prepared to fight for themselves, we cannot abandon them without disaster to peace and to our interests throughout the world.
>
> The reader may find it incomprehensible that Dean foresaw such dire consequences from the fall of South Vietnam, but I cannot overstate the impact our generation's experiences had on him (and, more or less, on all of us). We had lived through appeasement at Munich; years of military service during World War II fighting aggression in Europe and Asia; the Soviet takeover of Eastern Europe…(p. 195).

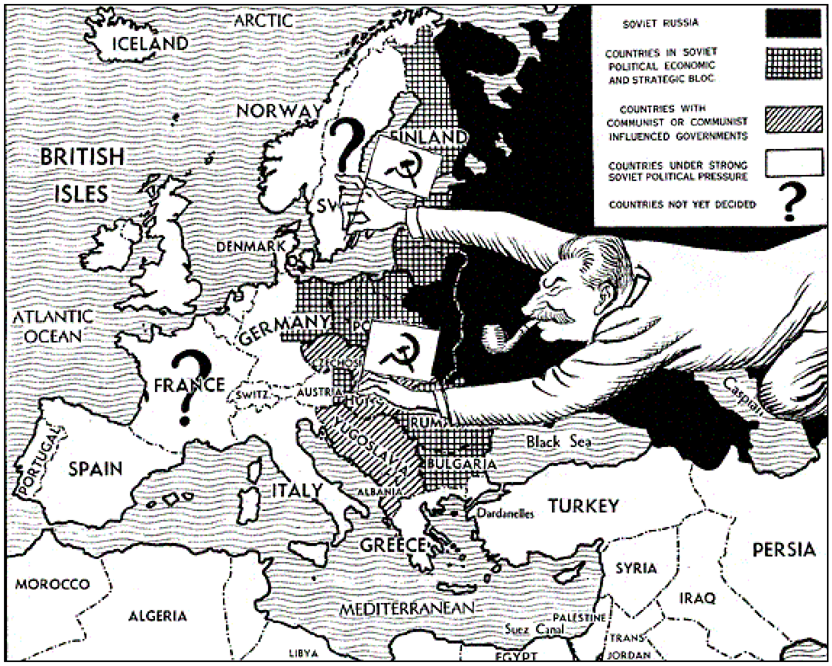


Figure III-1. Drawing by the British cartoonist Leslie Illingworth, published in June 1947, depicting the threatening reach of Stalin and the spread of the communist ideology in postwar Europe.

The second element of the postwar cultural situation was America's long history of antimilitarism. By antimilitarism, Edwards does not mean to imply pacifism or objection to armed force itself, but rather an "anti-power ethic" that strongly values limits on political power, hierarchy, and authority. As a result of their colonial experience with large garrisoning European armies, early generations of Americans understood both the importance of military power in international conflict and the dangers it posed in domestic political life. Prior to World War II, what American society strongly sought to avoid was not so much war itself as the permanent presence of a powerful national military institution (Figure III-2). However, the U.S. military success in World War II, the occupation of Germany and Japan, the smooth transition to the Cold War, and the U.S. emergence from the war relatively unscathed as a world power all contributed to a rapidly changing perception among the American populace of the need for a large military force. Additionally, technological factors such as atomic weapons and the maturation of air warfare created the possibility for a U.S. sphere of influence that extended well beyond North America. Consequently, the Cold War marked the first time

in its history that the U.S. maintained a vigorous military presence in peacetime. Even so, the longstanding American tradition of antimilitarism ensured that the form of this military force was different from the more traditional European and Soviet approaches that relied on large numbers of men under arms (Edwards, 1988, pp. 248–249).

Figure III-2. Typical antimilitarism cartoon from 1914 (source unknown).

The third element of the postwar political and cultural situation was the rise of science-based military power. At the end of World War II, science and engineering were widely viewed as being largely, in not entirely, responsible for the ultimate Allied victory. The crowning technological achievement of the war, the atomic bomb, represented nothing less than the military apotheosis of science. During the war, engineering academies like the Massachusetts Institute of Technology and the California Institute of Technology played major roles in the war effort, thereby increasing their base of political power and prestige. Their scientists and engineers, embracing the American tradition of pragmatic enterprise, showed that they could create impressive weapons

when given virtually unlimited resources. Consequently, the postwar period saw the emergence of a powerful and self-conscious science and engineering lobby and a permanent governmental association with science, largely mediated through the military services. The postwar scientific community therefore enjoyed an unprecedented sense of community, and its wartime miracles had won them patrons among the political and military leadership (Figure III-3) (Edwards, 1988, p. 249).[3]

Figure III-3. *Time* magazine cover from April 1957 celebrating two visionary scientist-engineers, Dean Woolridge and Simon Ramo, who are widely credited with introducing the high-technology, science-based, systems-oriented management approach (Hughes 1998).

According to Edwards, these three key elements of post-World War II U.S. political culture contributed to the sense of the world as a closed system accessible to American technological control. In particular, the postwar partitioning of Europe created the sense for most Americans that the world was now closed, being fully occupied by the

[3] Again, an illustration of another element of Edwards' thesis is provided in Robert McNamara's (1995) memoir. In recounting the members of the "Wise Men," an informal group called together by President Johnson to provide council on the Vietnam War, McNamara describes "distinguished Harvard chemist George Kistiakowsky" as personifying "the interrelationship of science and politics in the nuclear era" (p. 196).

apocalyptic struggle between the American and Soviet superpowers (Figure III-4). Moreover, under the Truman Doctrine and the Marshall Plan, the world had become a system to be protected and manipulated by the U.S. government. As a result, the demonstrated ability of science and engineering during World War II, and the global nature of the conflict with the Soviet Union, served to both justify and exaggerate the U.S. focus on high technology. The U.S. experience with the atomic bomb during World War II held out the promise of unlimited military power through American technological ingenuity. Simultaneously, the policy of containment required an ability to intervene with military force anywhere on the globe. In addition, the American tradition of antimilitarism further focused strategic planning on technological solutions as evidenced by the Strategic Air Command, which rose to prominence exactly because it required mainly money and equipment and not large numbers of troops (Edwards, 1988, pp. 250–251).

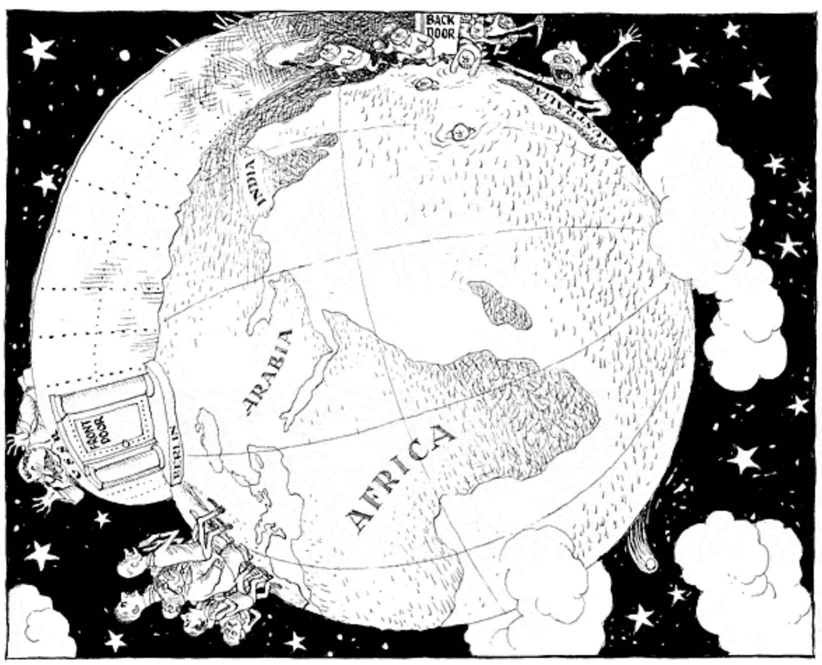


Figure III-4. Drawing by the British cartoonist Leslie Illingworth, published in May 1950, depicting a partitioned globe and emerging communist threats all over the world.

2. Technological Determinism

Thus, the rapidly evolving postwar geopolitical concerns of the U.S. as a world power shaped a strategic discourse centered on high technology. Such was the situation described by Edwards (1996):

> The primary weapons of the Cold War were ideologies, alliances, advisors, foreign aid, national prestige—and above and behind them all, the *juggernaut of high technology* [emphasis added]…Of all the technologies built to fight the Cold War, digital computers have become its most ubiquitous, and perhaps its most important, legacy. Yet few have realized the degree to which computers created the technological possibility of the Cold War and shaped its political atmosphere, and virtually no one has recognized how profoundly the Cold War shaped computer technology. Its politics became embedded in the machines—even, at times, in their technical design—while the machines helped make possible its politics (p. ix).

Implicit within this strategic discourse was the assumption that superior technology and weaponry would be a guarantor of combat success. Watts (1996), in his discussion of the basic relationships between doctrine, technology, and war within the domain of air warfare, describes this "hypothesis that technically superior hardware *often* or *always* guarantees success in combat [emphasis in original]" as "technological determinism" (p. 10).

So what is the problem with *technological determinism*? As noted by retired Major General Irving Holley (2004), a respected authority on military innovation, there is the unpleasant fact that technological determinism is not a sufficient cause for military success:

> …the thesis that superior arms favor victory, while essential, [is] insufficient unless the "superior arms" are accompanied by a military doctrine of strategic or tactical application that provides for full exploitation of the innovation. But even doctrine is inadequate without an organization to administer the tasks involved in selecting, testing, and evaluating "inventions." The history of weapons in the United States is filled with evidence on this point (p. 70).

Hence, Holley cautions that the hypothesis inherent in technological determinism is only potentially true when superior institutional weapon system acquisition practices yield

innovative technologies that are wed to thoughtful doctrine. So what can be said about the history of the U.S. Army from 1965–1985 in regards to Holley's concepts and ideas? Was a general institutional concern with technological superiority, if not outright technological determinism, a characteristic of the U.S. Army during this period? Answering such questions is of some importance in formulating a historical perspective of the emergence of the Army's MANPRINT program, which was the progenitor of the Defense Department's HSI program. Toward that end, it is necessary to consider the design and development of the Army of Excellence in the 1980s, which itself was the culmination of a massive tactical reorganization of the Army that created, in large part, the necessary preconditions for the historic emergence of the Army's MANPRINT program.

3. The Evolution of the Army of Excellence

In the midst of the first large scale troop reductions of the Vietnam War in 1969, the Nixon administration announced their "Guam Doctrine," which attempted to scale back the defense establishment with the objective of being able to fight a "1½" war contingency. This new doctrine was interpreted to mean that the Army should prepare to engage in a general war, likely in the European theater, and in a minor conflict, presumably a third world counterinsurgency. However, Nixon's vision for a smaller Army quickly faced growing challenges. U.S. intelligence agencies in the early 1970s observed that the Soviets were both modernizing and enlarging their armored forces in Europe and were stationing these forces ever further westward. Richard Stewart (2005) provides a somewhat more stark assessment of the perceived strategic reality of the time in his history, *The United States Army in a Global Era, 1917–2003*:

> If general war had come to Europe during the 1970s, the U.S. Army and its North Atlantic Treaty Organization (NATO) allies would have confronted Warsaw Pact armies that were both numerically and qualitatively superior. With the Army mired down in Vietnam, and with modernization postponed, this was a very sobering prospect (p. 377).

Additionally, the Arab-Israeli War in October 1973 was a watershed event for military planners. It vividly highlighted the increased battle tempo and materiel lethality of

modern war and called into question the Army's Vietnam-era concentration on infantry-airmobile warfare at the expense of other forces (Stewart, 2005):

> American observers who toured the battlefields of Egypt and Syria began to create a new tactical vocabulary when they reported on the "new lethality" of a Middle Eastern battlefield where in one month of fighting, the Israeli, Syrian, and Egyptian armies lost more tanks and artillery than the entire U.S Army, Europe, possessed. Improved technology in the form of antitank and antiaircraft guided missiles, much more sophisticated and accurate fire-control systems, and vastly improved tank cannons heralded a far more costly and lethal future for conventional warfare. […] It seemed clear that in future wars American forces would fight powerful and well-equipped armies with soldiers proficient in the use of extremely deadly weapons. Such fighting would consume large numbers of men and quantities of material. If became imperative for the Army to devise a way to win any future war quickly (pp. 377–378).

With many in the Army at the time concerned that they could not presently fight this type of new war, the Army set course on a decade of modernization and reform (Stewart, 2005, pp. 377–378).

John Romjue (1993), in writing the official history of the Army of Excellence for the U.S. Army Training and Doctrine Command (TRADOC) historical monograph series, calls out the "central importance" of "the personal push and stamp given to the Army's structural modernization and reform by Army Chiefs of Staff of the era, in particular General Edward C. Meyer (1979–1983) and General John A. Wickman, Jr. (1983–1987), as a well as by the early TRADOC commanders, General William E. DePuy (1973–1977), Donn A. Starry (1977–1981),…and William R. Richardson (1983–1986)" (p. 2). Accordingly, we will follow the story of the Army of Excellence largely through these "personal pushes and stamps." Much of this story comes from Romjue's excellent history on the subject.

It is worthwhile to take a moment to provide a short primer for those, like your author, who are not familiar with U.S. Army tactical organization. Since World War I, the basic ground unit in the Army, capable of sustained independent action, is the division. For that reason, the division has been the focus of tactical organization in the Army. Division structures are periodically redesigned as a result of their perceived

obsolescence in the face of anticipated conditions of future battle. Over the last century, each redesign has involved a progressively increasing application of technology to the division, due in part to the following: 1) the increasing mechanization of the fighting force, and 2) the extension of technology into virtually all the division's combat and support functions. One design of particular note to our story is the ROAD (for Reorganization Objective, Army Division) division, a 15,500-man structure consisting of a common division base and three maneuver brigade headquarters to which maneuver battalions—infantry, armored, mechanized infantry, airborne, or airmobile—were flexibly attached. The ROAD division was implemented between 1962 and 1964. Thus it was with the ROAD division that the Army went to war in Vietnam in 1965, and it was the ROAD division that formed the ground defense of Europe throughout the middle decades of the Cold War (Romjue, 1993, pp. 4–6).

With that background, we now introduce the first major protagonist of the story, General William E. DePuy, who from 1973 to 1977, was the first commander of the newly established U.S. Army TRADOC. General DePuy, an infantry officer in World War II and commander of the 1st Infantry Division in Vietnam, surveyed conditions on the modern battlefield and observed many of the same lessons that he and his men had learned painfully in World War II. Convinced that advances in weaponry were driving a tactical revolution in ground warfare that rendered the ROAD division obsolete, DePuy set in motion in 1976 a restructuring study of the heavy division. A major concern driving DePuy's thinking was that the volume and array of firepower now available to the company commander had exceeded manageable quantities. Additionally, he believed if the Army was to best exploit new weapons, organizational structures needed to be built around these new weapons rather than grafting new weapons onto existing organizational structures. Accordingly, the Division Restructuring Study (DRS) was carried out by TRADOC headquarters between May and July 1976 by a small group under DePuy. DePuy constrained the DRS to focus only on armored and mechanized infantry (i.e., heavy) divisions. The resulting proposed 17,800-man DRS heavy divisions featured significant changes to include smaller companies and smaller but more numerous maneuver battalions. The DRS heavy division was approved for testing in the 1st Calvary

Division at Fort Hood, Texas with favorable results, but ultimately, the DRS heavy division did not survive. Doubts arose in the Army Staff and elsewhere about the smaller units, the brigades' increased span of control, and other features. When DePuy's successor, General Donn A. Starry assumed command at TRADOC in July 1977, he expressed doubts that essentially sealed the demise of the DRS heavy division (Romjue, 1993, p. 8).

General Starry, a noted cavalry leader in Vietnam and a soldier-scholar, arrived at TRADOC straight from command of the V Corps in Germany, where he had the opportunity to develop a firsthand appreciation of the Soviet's overwhelming forces. Under Starry, TRADOC began a comprehensive reorganizational effort, Army 86, which continued and extended the aim of the Division Restructuring Study work. This effort was initiated with the Division 86 Study in August 1978, which, like the DRS, focused on the heavy division—the element of the fighting Army critical to the primary strategic theater in central Europe. Unlike his predecessor, General DePuy, who structured his DRS heavy division specifically upon new weapon systems, General Starry took a systems engineering approach to the division problem (Romjue, 1993):

> Starry's whole approach was 'a systematic breakdown into the division's specific tasks and subfunctions and then a reconstruction into a coherent whole or division capability.' What he wanted division designers to do was to leave behind parochial branch approaches to battles and to see their challenge instead in terms of the major functions that he believed characterized modern battle (p. 9).

Out of his V Corps experience and his functional vision came the concept of "seeing deep" to the enemy's follow-on echelons, which led to a doctrinal focus on deep attack to disrupt the enemy's second echelon forces—what eventually became known as AirLand Battle (Romjue, 1993, p. 9).

Having just mentioned AirLand Battle, we now consider a parallel thread in this story—one involving the same protagonists, but focused instead on the evolution of corresponding operational concepts. In the immediate post-Vietnam era, the emphasis of Army planning refocused on large scale conventional war in Europe (Stewart, 2005):

> Generals Abrams and DePuy and like-minded officers believed the greatest hazard, if not the greatest probability of war, existed [in Europe]. They conceived of an intense armored battle, reminiscent of World War II, to be fought in the European theater. If the Army could fight the most intense battle possible, some argued, it also had the ability to fight wars of lesser magnitude (p. 387).

This focus on conventional war in Europe necessitated a change from the doctrine that prevailed during the middle decades of the protracted Cold War. A new operations field manual is the method by which the Army promulgates and codifies its latest doctrine. General DePuy began a post-Vietnam doctrinal renaissance by rewriting much of the 1976 edition of Field Manual (FM) 100-5, *Operations*, the Army's central doctrinal publication. DePuy's FM 100-5 touted the concept of Active Defense, which once more focused on the primacy of defense. It emphasized the importance of the tank as the pivotal element of land forces, promoted concentration of fires rather than forces, and advocated for the replacement of tactical reserves with the lateral movement of unengaged units behind strong covering fires. Such a radical departure from earlier doctrine proved both controversial and difficult to implement at the time, leading to an extended doctrinal and tactical discussion in the service journals that served to clarify and occasionally modify the manual (Stewart, 2005, pp. 378–379).

When General Starry succeeded DePuy at TRADOC, he directed a substantial revision of FM 100-5 to concentrate on the offensive and stress aggressive operations in depth with an increased emphasis on the exploitation of tactical air power—a concept that became known as AirLand Battle doctrine. The major shift in Army doctrine was officially signaled by the publication of the 1982 edition of FM 100-5, which documented the changeover from Active Defense to AirLand Battle. The manual stressed that the Army had to "fight outnumbered and win" the first battle of the next war, an imperative that, in turn, required a trained and ready peacetime force. It acknowledged the preeminence of the armored battle in warfare and the tank as the single most important weapon in the Army's arsenal. The manual also embraced the traditional concepts of maneuver warfare as the means for achieving success on the field of battle. More importantly, as described by Stewart (2005), AirLand Battle doctrine:

> …explicitly acknowledged the growth of technology both as a threat and as a requirement for new equipment to meet the threat. The U.S. Army and its NATO allies could not hope to match Soviet and Warsaw Pact forces either in masses of manpower or in floods of materials. To that extent, AirLand Battle served as a basis for both an organizational strategy and a procurement rationale. To fight outnumbered and survive, the Army needed to better employ the nation's qualitative edge in technology (p. 379).

Thus, AirLand Battle doctrine proved useful to the Army because it helped both define the proper weapon systems for its execution and the appropriate organization of military units for battle (Stewart, 2005, p. 379).

The primary justification for technologically superior weapons came from the military theorists of the time who generally believed that a defending army could reasonably expect success if an attacking army had no greater than a 3:1 advantage in combat power. The problem for the U.S Army was that the best intelligence estimates in the 1970s gave the Soviets an advantage that was significantly greater than 3:1. Moreover, continuing budget constraints made the option of increasing the size of the U.S. military to match Soviet growth untenable. Consequently, the Army looked to solve this problem by relying on superior technology that, it was hoped, would allow the Army to defeat an enemy at ratios higher than 1:3. To that end, in the early 1970s, the Army began work on its "big five" weapon systems: a new tank (the M1 Abrams tank), a new infantry combat vehicle (the M2 Bradley infantry fighting vehicle), a new attack helicopter (the AH-64A Apache attack helicopter), a new transport helicopter (the UH-60A Black Hawk utility helicopter), and a new antiaircraft missile (the Patriot air defense missile). However, these were by no means the only significant equipment modernization programs. Other important Army procurements included the multiple launch rocket system; a new generation of tube artillery to upgrade fire support; improved small arms; tactical-wheeled vehicles such as a new 5-ton truck and utility vehicle (the high-mobility multipurpose wheeled vehicle, or HMMWV); and a family of new command, control, communications, and intelligence (C3I) systems (Stewart, 2005, pp. 379–384).

The coincidence of several factors had significant effects on the design of these new weapon systems. Stewart (2005) describes some of the more important factors:

> Among the most important was the flourishing technology encouraged by the pure and applied research associated with the space programs. Although the big five [weapon systems] originated in the years before AirLand Battle was first enunciated, that doctrine quickly had its effect on design criteria. Other factors were speed, survivability, and good communications, essential to economize on small forces and give them the advantages they required to defeat larger, but presumably more ponderous, enemies. Target acquisition and fire control were equally important since the success of a numerically inferior force depended heavily on the ability to score first-hit rounds (p. 380).

Despite the clear future vision provided by AirLand Battle, the complexity of the space age technologies and the conflicting nature of many of the doctrinally relevant design criteria made it difficult for the Army to bring system concepts to fruition (Stewart, 2005):

> Even such simply stated criteria were not easy to achieve, with compromises and trade-offs often necessary between weight, speed, and survivability. All of the weapon programs suffered through years of mounting costs and production delays. A debate that was at once philosophical and fiscal raged around the new [weapon systems], with some critics preferring simpler and cheaper [systems] fielded in greater quantities. The Department of Defense persevered, however, in its preference for technologically superior systems and managed to retain funding for most of the proposed new weapons. Weapon systems were expensive, but defense analysts recognized that personnel costs were even higher and pointed out that the services could not afford the manpower to operate increased numbers of simpler weapons. Nevertheless, spectacular procurement failures, such as the Sergeant York Division Air Defense (DIVAD) weapon, kept the issue before the public; such cases kept program funding for other equally complex weapons on the debate agenda (p. 380).

Nevertheless, a close relationship between doctrine and technology swiftly developed. Weapon system modernization led doctrinal thinkers to consider even more ambitious concepts that would exploit the potential capabilities the new systems promised (Stewart, 2005, p. 385).

While General Starry directed development of doctrinal concepts that would take advantage of the increased combat power of the new materiel systems that were becoming available, he also focused on designing the organizations that could exploit them—and hence, the segue back to our earlier story of the Division 86 Study. The method of the Division 86 Study departed from that of DePuy's small study-cell approach used in his Division Restructuring Study. The Division 86 Study was a major year-long enterprise involving several task forces at selected Army schools and employing analysis and war gaming of alternative unit structures—Romjue (1993) suggests that "its depth may have been unprecedented in Army tactical unit reorganization" (p. 10).

The resulting Division 86 heavy division, much of the structure of which survived into the 1980s Army, numbered approximately 20,000 soldiers. There were six tank battalions and four mechanized infantry battalions in its armor version, and five tank and five mechanized infantry battalions in its mechanized infantry form. It also added a significant new component in an air cavalry attack brigade as well as expanded the division artillery (Romjue, 1993, p. 10). The new brigade support battalions of Division 86 implemented the concept of "arm, fuel, fix, and feed forward." Additionally, as described by Romjue (1993):

> An important design element was the building into the heavy division of what planners called 'R3': personnel strength providing robustness, redundancy, and resilience for critical division control functions and key combat tasks. The heavy divisions in Europe facing the overwhelming might of the Warsaw Pact forces had to be heavy and then some (p. 9).

Collectively:

> …the Division 86 organizations were keyed to concepts of maximum firepower forward; improved command control; increased fire support, air defense, and ammunition resupply; and an improved combining of arms. The structure imposed an increased leader-to-led ratio, with smaller and less complex fighting companies and platoons (p. 10).

The logic behind the Division 86 design was clear: 1) to fight and win on a conventional, high-intensity battlefield in Europe without relying on tactical nuclear weapons, and 2) to

field forces that vastly increased the depth over which the enemy could be attacked (Hawkins & Carafano, 1997).

General Edward C. Meyer, the Chief of Staff of the Army, approved the Division 86 design in principle in October 1979 and approved it for implementation in decisions of August and September 1980 (Romjue, 1993, p. 10). However, his 1980 decision carried significant future manpower costs (Romjue, 1993):

> In the defense climate of 1980, Army force design focused on the serious threat posed by the massive Soviet buildup. That concern, and not end-strength Army totals, dictated the initially strong designs of [Division] 86. The election to the U.S. presidency in the fall of 1980 of Ronald Reagan, a strong defense advocate, might have been expected to provide the needed Army manpower increases. Reagan was strongly committed to an accelerated buildup of American military power to enable the nation to meet the Soviet challenge in Europe and elsewhere. His accession did indeed soon lead to increased budget commitments. In that general trend, however, and as planning began toward conversion to the new heavy division designs, the Department of the Army did not move to press for the significantly higher active-component end-strength needed to accommodate the larger Division 86 designs (p.13).

Repeated attempts by the Army in the early 1980s to raise the manpower ceiling by 5,000 to 15,000 men did not succeed at either the level of the Defense Department or Congress (Romjue, 1993, p. 21). In the meantime, the modernization of the force was proceeding apace. In the latter half of 1981, Department of the Army and TRADOC planners began to examine alternative solutions to the manpower problem. In November 1981, the Department of the Army select committee, chaired by the Vice Chief of Staff of the Army, General John W. Vessey, Jr., convened to take up the problems of Division 86 transition. The select committee, recognizing that the Division 86 design was not affordable with the Army end-strength levels established through 1988, directed TRADOC to reduce the heavy divisions by ten percent to 18,000 soldiers. The subsequent Division 86 Restructuring Study, carried out by TRADOC, attempted to downward structure the heavy division while keeping the basic design intact with combat power undiminished. In March 1982, the Army Chief of Staff decided on a division

reduced not by 2,000 but by 1,000 from the original 20,000 structure. Nevertheless, the 1982 restructuring exercise ultimately failed to materially affect the manpower impasse (Romjue, 1993, pp. 13–15).

It is time to consider yet another thread in this story—one that includes, if only tangentially, a future organizational sponsor of MANPRINT, namely Lieutenant General Robert M. Elton. Recall that the Division 86 study was but one of several Army 86 elements. Starry's Division 86 study was driven, in large part, by a shift in U.S. national military strategy in 1978, which implemented, in conjunction with NATO, a conventional force buildup in Europe to match the Warsaw Pact (Hawkins & Carafano, 1997). Similar studies, collectively known as the Army 86 Studies, considered the correct structure for the infantry division, the corps, and larger organizations. One of these studies, Infantry Division 86, was begun in 1979 and reflected another transition in the national military strategy: the U.S was broadening its focus again beyond Europe to consider regional contingency missions. Up to the close of the 1970s, U.S. national military strategy paid little attention to the prospect of military action elsewhere in the world other than Europe, leading the Army to focus almost exclusively on the development of heavy forces. It was only in 1979, with the Soviet invasion of Afghanistan and the Iranian hostage crisis, that senior policy makers began seriously considering the need for flexible contingency forces including more rapidly deployable light divisions (Romjue, 1993, pp. 10–13, 15).

As late as 1979, Army plans called for mechanizing all the remaining standard infantry divisions, exclusive of the 82nd and 101st Airborne. However, in that same year, General Edward C. Meyer, a cavalry leader in Vietnam and an advocate of lightness, ascended to the Army Chief of Staff position and quickly took steps to stop the mechanizing trend. General Meyer believed there was another way, other than "heavying up" (i.e., mechanization), to make the standard infantry divisions effective: increased technology. Meyer convinced then Secretary of Defense Harold Brown to forego a plan to mechanize the 9th Infantry Division, proposing instead that it be redesigned to obtain many of the characteristics of a heavy division through innovative organization and new technology. The search for a light division design in 1979 took two courses, though no such separation of effort was originally planned. In late 1979, the Army 86 planners

began the Infantry Division 86, or ID 86, Study (Romjue, 1993, pp. 15–16). Generals Meyer and Starry developed a detailed concept for ID 86 (Hawkins & Carafano, 1997):

> It had to be able to conduct worldwide contingency operations as well as deploy rapidly to reinforce forward NATO forces. To do this the division would need increased mobility, flexibility and firepower. General Meyer detailed two design constraints. The division would be capped at 14,000 soldiers and limited to equipment that could deploy in C-141 aircraft. The designers would have to depend on advanced technologies to enable these smaller divisions to accomplish their diverse and demanding missions.

This dual concept of a nonmechanized light division that could be effective as a rapid deployment division in third world contingencies as well as on the armor dominated battlefield of Europe proved a constant frustration for planners. The ID 86 Study conducted during 1979–1980 excluded tank and mechanized infantry battalions, but emphasized a strong antiarmor capability, hopefully provided by "high technology." However, in the end, the NATO half of the infantry division's dual mission drove an ID 86 design that was not "light" in men, equipment, or support. Plans went forward to test the resulting 18,000-man ID 86 design using the 9th Infantry Division at Fort Lewis, Washington as a so-called "high technology test bed" for transition (Romjue, 1993, p. 16).

The High Technology Test Bed, or HTTB, was the second developmental course spawned by the ID 86 Study. Though not initially viewed by the Army 86 planners as a separate effort, it in fact evolved in that direction. By official agreement in October 1980, the HTTB was the united effort of TRADOC, the Army Materiel Development and Readiness Command, and the Army Forces Command. At the direction of the Department of Army, the 9th Infantry Division Commander (i.e., Lt. General Elton) was the HTTB test director (Romjue, 1993, p. 17). As is often the case with collaborative agreements, differing perceptions soon developed between TRADOC and the 9th Division as to the relative relationship between technology and organizational design (Romjue, 1993):

> Was the test bed to test the Infantry Division 86 concepts and organizations and infuse new high technology systems into the 9th Division, as TRADOC understood? Or was the focus first on the infusion

> of new technology and on innovative and enhanced deployability unhampered by conceptual structures—the 9th Division's understanding of things? The upshot of the disagreement—the decision by General Meyer in April 1981 that ID 86 was the starting point only—effectively set the 9th Infantry Division test bed upon the effectively independent track it subsequently pursued under Meyer to develop high technology light division designs and ideas (p. 17).

Thus, the high technology light division subsumed the ID 86 effort to become the focus of light infantry division design—technology would drive organizational design. In the end, however, no high technology light division eventuated from the test bed. A major reason for this failure was that the weapons programs on which the concept depended failed to gain funding. Chiefly involved were light or "fast attack" vehicles and armored assault gun vehicles (Romjue, 1993, p. 17). In summary judgment, the Army 86 work failed to realize a design for the Army's main light force element (Romjue, 1993):

> In 1982–1983, Army force designers found themselves no farther along toward a new realistic infantry division design than they had been four years earlier. High technology testing had not proved sufficiently convincing to pose the "high-tech" route as an answer (p. 20).

The year 1983 saw the onset of what Army planners called the "bow wave" of force modernization as new weapons and equipment were fielded in earnest to the divisions in U.S. Army, Europe. The accession of General John A. Wickham, Jr., to the post of Chief of Staff of the Army in June 1983 set in motion a major new design and structuring approach to the Army's tactical units—the Army of Excellence (AOE)—that effectively superseded the Army 86 design and modernization effort. As early as April 1983, while still Vice Chief of Staff of the Army, General Wickham formed a small group of officers under Brigadier General Colin Powell, called "Project 14," to identify issues he expected to face. Among the findings of the Project 14 team were the need to move in the direction of more light infantry and the common recognition that Division 86 was not affordable. During this period, General Wickham notified General William R. Richardson, who had taken command of TRADOC in March 1983, that he wanted TRADOC to develop a light division of 10,000 personnel. General Richardson, who was involved in a major portion of the Army 86 force design, agreed but advised the Chief of Staff that such a redesign should be part of a larger whole. Richardson's thinking was to

line up the Army by its several corps and by elements (i.e., combat, combat support, and combat service support) and to design and structure it in a way that the light infantry divisions would best fit in (Romjue, 1993, pp. 23–24).

General Wickham, looking ahead to TRADOC's future concept, AirLand Battle 2000, believed the Army needed to move with reasonable urgency toward a lighter force design. He was also interested in not only preserving, but actually increasing combat strength (Romjue, 1993, p. 25). Wickham looked to history—relatively recent history at that—for the way forward (Romjue, 1993):

> Ten years earlier when the Army, withdrawing from Vietnam, had been reduced to a low of thirteen divisions, the Army Chief of Staff General Creighton Abrams, eyeing the rising Soviet threat to NATO Europe, had set a goal of 16 Active Army divisions by 1976 without Army end-strength increases. Abram's initiative, which had been carried through to completion after his untimely death in office in September 1974, had achieved that goal through a paring-back of the support structure and employment of reserve component "roundout" brigades and other units for the Active Army divisions. What that meant was that some active divisions commanded only two active brigades, filling out their strength with a reserve unit as the third brigade. Those measures were strongly supported by Secretary of Defense James Schlesinger. Not only did they convert fat to muscle in terms of combat units and anchor the Army's future war fighting commitment in its reserves as well as in its active forces; the Abrams initiatives also sent a deterrence message (p. 25).

General Wickham elected to resurrect and employ the Abrams paradigm. Facing the reality of an inflexibly capped Army end-strength and the twin dilemmas of a continuing Soviet threat in Europe and a growing need for light, rapidly deployable contingency forces to meet third world crises, Wickham pushed for a force design initiative that placed a premium on replacing support strength with combat units (Romjue, 1993, pp. 24–25).

In preparation for the 1983 Summer Army Commanders' Conference, the major Army commands began surfacing issues under the theme of "resources for excellence" (Romjue, 1993, pp. 28–29). Among these issues was the ongoing work by TRADOC in

assisting the Army to field and transition to the organizations of Army 86, and the necessity to deal with the force structure dilemma arising from the Army 86 designs (Romjue, 1993):

> The specifics of the dilemma were that, in order to fulfill the organizational designs of Army 86, the Army's projected active force structure would have to increase to 836,000 personnel in the coming decade. That manpower total exceeded considerably the 780,000 end-strength imposed by foreseeable budgetary constraints. Given that limitation, and the assumption that none of the Active Army divisions would be inactivated, TRADOC needed to describe how to modify the Army 86 force structure to conform to the end-strength reality. [It was] advised that the following steps would be necessary: further reduce the heavy division; suggest design options for smaller light divisions; examine the design of the special operations forces; and consider new support ratios between division, corps, and echelons above corps (p. 29).

Accordingly, General Richardson became interested in the disproportionate growth in combat support and combat service support in recent years at the expense of the combat elements of the force structure. The trend had begun with the increase in 20,000 spaces of what Army planners called the division force equivalent, or DFE, a planning term referring to the division plus those nondivision forces needed to support it in combat. Both Division 86 and the high technology light division had a direct bearing on this trend as they reduced the infantry structure and increased support (Romjue, 1993, pp. 29–30).

During the Army Commanders' Conference of August 1983, General Wickham tasked TRADOC to develop a total force design that fully considered the factors of supportability, deployability, threat, and manpower ceiling. General Richardson directed the Fort Leavenworth Combined Arms Center to form an AOE study group and provide recommendations to the Chief of Staff by the Army Commanders' Conference of October 1983 (Romjue, 1993, pp. 35–37). The crux of the study was the question, "How is the Army going to pay the manpower bill?" (Wild, 1987, p. 5). The Army Staff also provided the following points of guidance for TRADOC's AOE study (Dupay, 1988, p. 6):

1) The recommended designs would not exceed the Army's programmed end-strength.

2) Determine whether the Army could be manned at Authorization Level of Organization 2, which equated to manning units at no less than 90 percent of required wartime strength.
3) Develop a proposal for a light, division-size force for rapid deployment for contingency missions.
4) Recommend reductions to the end-strength of heavy divisions to increase maneuverability of organizations.
5) Redesign corps and echelon above corps to improve their combat capability.

In sum, the primary objective of the AOE study was to address the "hollowness" and lack of strategic deployability that had resulted from the Army 86 designs (Dupay, 1988, pp. 4–5).

The AOE study methodology began with the design of the light infantry division. Design criteria, in addition to the manpower limitation of approximately 10,000 soldiers, included the following (Wild, 1987, pp. 5–6):

1) The division force design would be optimized for employment at the lower end of the conflict spectrum in a contingency mission, yet retain utility for employment at higher conflict levels as might be anticipated in Europe.
2) The division would be deployable in no more than 500 C-141 sorties.
3) The division would contain approximately 50 percent infantry.
4) The division design would have nine maneuver battalions.

Heavy division redesign followed with the goal of retaining the combat capability of the Division 86 design while reducing the division end-strength. AOE modifications moved some functions out of the heavy division to the corps or higher and reduced personnel robustness, redundancy, and resilience (R3) from the remaining functions with the goal of achieving economies through centralization (Dupay, 1988, pp. 18–20). Such reductions could not be made without some loss in capability, but wherever possible, cuts were made primarily in support and service support functions (Wild, 1987, p. 6).

Lieutenant Colonel Arthur Dupay, in a 1988 Army War College study of the AOE, asserts that the primary objective of the AOE was to make combat service support functions (the "tail") the primary bill payers for two new light infantry divisions (the

"teeth") (p. 9). In so doing, the Army could increase its overall "tooth-to-tail" ratio in accordance with the Abrams paradigm. Additionally, by developing the new light infantry divisions such that they were greatly reduced in size and revised in concept from existing and proposed designs—that is, General Wickham's goal of a 10,000-man division—the spaces saved could be applied to other changes needed such as the full manning of Active Army units (Romjue, 1993, p. 30).

Changes to the tooth-to-tail ratio were guided, in large part, by the Logistics Unit Productivity Study, or LUPS (Romjue, 1993, pp. 49–50). Realizing that many of the formulas for combat service support force requirements were based upon the Army's experience in World War II, the Army Logistics Center conducted the LUPS in 1982 to examine "ways to replace as many soldiers in combat service support units as possible with modern high-technology equipment, seeking efficiencies from productivity enhancement" (Wild, 1987, p. 6). According to Dupay (1988):

> The key was to improve the reliability, availability, maintainability, and durability of equipment; reduce weight, volume and manpower requirements; and improve logistics unit and systems productivity and throughput. Specific issues such as the palletized loading system (PLS), automated pipeline construction system, robotic refueling systems, and expert diagnostic systems were just some of the reasons the logistics community assumed the "can do" attitude and handed over more than 15,000 spaces for the AEO initiative (p. 24).

Overall, LUPS was credited with freeing upwards of 15,000 combat service support soldiers for reassignment, thereby facilitating the creation of the two light infantry divisions (Wild, 1987, p. 7). However, Dupay (1988) notes that these projected manpower savings were based on efficiencies through technologies that were still early in the research, development, test, and evaluation process (p. 24).

Despite the focus on AOE logistical changes, it was the creation of the light infantry division that was the real centerpiece of the AOE reorganization. The AOE light infantry division represented a significant break in the history of Army tactical organization. The light infantry division was fashioned for use primarily to respond to contingencies in the third world with only a collateral mission to reinforce heavy forces, and only then when terrain and circumstances were appropriate. The latter represented a

significant relaxation of the dual mission requirements that stymied the Army 86 planners working on the original high technology, light infantry concept (Romjue, 1993, pp. 45–48). Nevertheless, the design of the light infantry division was still dependent on some efficiency gained through technology (Dupay, 1988):

> The [light] division is composed primarily of fighters equipped with lightweight weapons systems [sic] which are supposed to be sustained by an austere support structure. The division was designed to capitalize on technological advances to enhance its performance and reduce the manpower required to perform essential battlefield tasks (p. 31).

Combat service support was limited to the minimum essential assets required for operations in contingency areas (Dupay, 1988, p. 31). As envisioned by General Wickham, the light infantry divisions were developed as a hard-hitting, elite force derivative of the Rangers. A premium was placed on the capabilities of the individual light infantry solider and his unit (Romjue, 1993, p. 53).

In far reaching decisions during October and November 1983, General Wickham endorsed the AOE design for planning and then implementation. He believed the AOE design combined affordability, high combat readiness, and strategic deployability and struck a sound balance between heavy and light forces (Romjue, 1993, p. 55). Another key figure in the MANPRINT story, General Maxwell R. Thurman, Vice Chief of Staff of the Army, also strongly supported the AOE design (Romjue, 1993, p. 38). On 10 January 1984, the Department of the Army issued further implementing decisions and instructions. The phased restructuring of the Army began in late fiscal year (FY) 1984 and extended throughout the next several years. Two active-component infantry divisions, the 7th at Fort Ord, California, to transition between late FY 1984 and late 1986, and the 25th at Fort Drum, New York, to transition subsequently, would convert to the new light infantry design (Romjue, 1993, pp. 52–56).

The story of the AOE continues, but it was the efforts through 1983 that culminated in the approved organizations of the Army of the 1980s. And so it is here that we will conclude this thread of the MANPRINT story. No major reorganization can escape controversy, and to a degree, the AOE is open to criticism that it overemphasized

combat power at the expense of support units. What should be appreciated is that the AOE design involved significant tradeoffs:

> In 1968, the Active Army had consisted on 18 $^{2}/_{3}$ divisions in an active force of 1.5 million personnel. In 1986, the Active Army's 18 divisions were carved from an end-strength of 780,000…The fielding of 18 divisions from so small a force had been achieved only by drastic cutbacks in combat support and combat service support in the active force and by the maintenance or placement of much of the support force…in the nonexistent "component 4" category. There was some degree of validity to the hollowness charge. But in no army in a democracy in peacetime will a fully adequate force be funded. If the Army of Excellence was not the best possible Army, it was an Army of the best affordable divisions and corps at the time. […] By maximizing combat power in more divisions but with no added Active Army end-strength, the AOE decisions left many corps and theater functions unmanned and some U.S.-based divisions dependent on less-ready reserve roundout brigades. That inadequacy was the price and prudent risk of General Wickham's decision, a decision supported by the Joint Chiefs of Staff, for the deterrence value believed to be gained. Facing worldwide defense challenges in the 1980s, the U.S. Army leadership chose more divisions and battalions, more forward combat strength and combat diversity, over the security of a force of fewer divisions, stronger in support, manned adequately top to bottom (Romjue, 1993, p. 126).

It is not my intent here to critique those tradeoffs, which is a task best left to a scholar of military history. What is important is that we gain at least an elementary cognizance of the doctrinal and organizational context within which the U.S. Army of the early 1980s worked—the Army from which MANPRINT emerged:

> The Army of Excellence was an Army built upon dilemmas rooted in the political and strategic currents of the early 1980s. Those omnipresent realities—a powerful and dangerous Soviet adversary, a global defense mission, an ongoing major cycle of weapon modernization, and an inflexibly capped Army end-strength too small for the force needed—were factors forcing Army leaders to a compromise of balanced heavy and light organizational designs. These designs were unavoidably imperfect yet remarkably sufficient for the historically unprecedented strategic challenge and responsibility faced and borne by the United States in the world-changing decade of the 1980s (Romjue, 1993, p. xiii).

Thus, Romjue echoes Edwards (see Section III-B1) by stressing the necessity of understanding the larger post-World War II U.S. political culture—the "omnipresent

realities"—when analyzing the Army's choice of military structure. For this very reason we will now turn, if only briefly, to the military reform drive that was playing out in the press and within the hearing rooms of Congress and which reached its zenith in 1983.

4. The Military Reform Movement

The period spanning Jimmy Carter's presidency (1977–1981) and Ronald Reagan's first term (1981–1985) was one of significant turmoil within the U.S. defense establishment. As described in Grant Hammond's 2001 book on John Boyd:

> The Vietnam War had ended with the fall of Saigon in April 1975. There was confusion about national security strategy, national military strategy, the transition from conscription to an all-volunteer force, military relevant technologies, arms control, budget battles for defense versus other needs, fights among the services on individual weapon systems and just why the war had been lost. After being lied to about Vietnam, the American public was increasing skeptical of its government's claims. In part, it was this uncertainty and confusion that gave rise to the opportunity for a military reform movement (p. 106).

During this period, the U.S. military had to reinvent itself after defeat and begin preparing for challenges on both ends of a spectrum of conflict that ranged from irregular warfare to strategic nuclear war. Simultaneously, there was a saga of public controversies surrounding force structure planning and nonperforming weapon systems (Hammond, 2001, p. 102). Robert Coram paints a succinct, if dire, picture of the state of affairs at the time in his 2002 biography of John Boyd:

> By 1978, both officers and enlisted personnel were leaving the military services in large numbers. They left not because of pay, as military leaders had said for the past few years, but because they were displeased with what they saw as a lack of integrity among their leaders. They thought careerism inhibited professionalism in the officer corps. The military also was having readiness problems; expensive and highly complex weapon systems were fielded before being fully tested. These systems were not only expensive to buy but expensive to maintain, and they rarely performed as advertised. Stories began to appear in the media of America's "hollow military." […] The military's answer was to place more emphasis on what it called the "electronic battlefield" by buying even more expensive and more high-tech weapons. Somewhere in the military there must have been those who sensed the system was headed towards a meltdown. If so, no one stepped forward to change it (p. 345).

In the wake of Vietnam, it became increasingly clear to those who sought to change the way the U.S. defense establishment worked that something more substantial was needed besides attacks against individual weapon systems, policies, or funding decisions (Figure III-5). These people, which included military and civilian Pentagon insiders, journalists, academicians, and members of Congress, coalesced to form a loosely affiliated group, known as the military reform movement, which became increasingly vocal and active in taking on the Defense Department (Hammond, 2001, pp. 101–113).

Figure III-5. Drawing by Ben Sargent published in September 1981, referencing the budget debate in the U.S. Congress. President Reagan had managed to get major cuts in social security, public service, and welfare prior to that. However, given rising unemployment, people and Congress were increasingly reticent to accept further cuts while the Defense Department was to get their usual funding. Defense is drawn here as an overly obese boxer who is already unable to get up, yet alone fight. The fairly small physique of the general behind him illustrates the fact that the Defense Department leaders had already lost control over the military forces' quantity. The next year, Congress rejected Reagan's plan for more cutbacks and forced a scaling-down of the defense budget (Fischer & Fischer, 1999).

Several different groups and many different issues came to be involved in the military reform movement. However, Hammond (2001) characterizes the main debate as being between technologists and reformers, although he acknowledges that this is

somewhat of an oversimplification that distorts reality (pp. 107–109). Hammond suggests that the best short explanation of the two schools of thought is that provided by Serge Herzog (1994):

> Succinctly stated, reformers hold the following positions: (1) overemphasis on high technology has driven cost of modern weapons out of control; (2) high technology has introduced a level of complexity that seriously hampers force readiness; (3) high technology is pushed in areas often irrelevant to success in combat and may even endanger its user; (4) the added increment in performance resulting from high technology rarely justifies the costs involved; and (5) high technology stretches acquisition and maturation, causing critical delays in technology integration and frequently unexplained technical problems (pp. 3–4).

In opposition to the reformers were those characterized as technologists who held the following views: "(1) technology acts as a force multiplier; (2) technology provides force flexibility; (3) technology has the potential to improve cost and equipment reliability and maintainability; and (4) technology is indispensible given the alternatives" (Herzog, 1994, pp. 6–7).

While reformers raised a broad set of defense issues beyond the high-technology weapons debate, key members of the reform movement were relatively unanimous in their criticism of *complex* high technology weapons. The critical question for reformers was whether high technology weapons inevitably led to too much complexity. Their specific argument rested, in part, on a "general relationship" between weapons complexity and low combat readiness. They suggested that increasing weapons complexity multiplies reliability and maintainability problems, thereby increasing ownership costs, particularly maintenance costs. They also argued that increasing complexity reduces combat force size (i.e., the tooth-to-tail ratio), supplies, spares, and munitions to inadequate numbers, necessarily resulting in a less capable force. Using tactical aviation as an example, reformers attacked the supportability of high technology weapons, claiming that an overemphasis on complexity resulted from an unsatisfactory approach to system design—an approach that emphasizes performance, schedule, and acquisition cost while ignoring the "ownership considerations" of logistic support, human

factors, and quality assurance.[4] Reformers advocated for weapons that were cheaper, less complex (but not devoid of new technology), easier to maintain, more autonomous, and facilitative of greater ease of training. Figure III-6 presents a chart from the Spinney Report[5] summarizing the reformers' case. The reformers believed the answer to the critical question in Figure III-6 was a resounding "no" (Kross, 1985, pp. 17–18, 57–59).

INCREASING WEAPONS COMPLEXITY REDUCES COMBAT READINESS

- Degrades combat skills by causing inadequate and unrealistic training.
- Increases reliability and maintainability problems.
- Increases cost of maintenance
- Increases dependence on large vulnerable support base.
- Increases economic inefficiency of plans.
- Slows modernization by increasing development/procurement lead times.
- Multiplies magnitude and likelihood of disaster.
- Increases vulnerability to countermeasures.
- Cuts forces, supplies, and munitions to inadequate numbers.

QUESTION

Do the distinctive characteristics generated by weapons *complexity* compensate for these negative qualities?

Figure III-6. Summary arguments against complexity from the Spinney Report (From Spinney, 1985).

[4] The focus on "ownership considerations" is a major thrust of current Defense Department policy guidance on HSI. Specifically, Department of Defense Instruction 5000.02 (2008) requires program managers to take into account total ownership costs and accommodate the user population that will operate, maintain, and support the system. Thus, we see elements of the military reform debate that were carried forward into Defense Department policy guidance on HSI.

[5] In making their case, the reformers relied on three major briefings – each unpublished, continually refined, and constantly updated. Virtually all arguments made by the reformers on high technology weapons and other subjects related to military reform can be traced back to these briefings. One of these briefings was Franklin C. Spinney's "Defense Facts of Life," a 4-hour presentation that was the reformers' definitive statement on the penalties of pursuing complex, high technology weapons in an environment of limited defense spending. This briefing is often called the *Spinney Report* (Kross, 1985).

The issues were, and still are, complex, but it is fair to say that the two sides of the reform debate differed less in their ends and far more in the means to accomplish them (Hammond, 2001, p. 108). One could argue that the underlying debate was really over the veracity of the hypothesis of technological determinism that undergirded defense acquisitions during the Cold War. This point is perhaps best illustrated by Hammond's (2001) characterization of the philosophy of John Boyd, the legendary maverick and military strategist at the center of the debate on the side of the reformers:

> Yet Boyd and the reformers understood a central reality of national security. It flowed from Boyd's affinity for trade-off analysis and his propensity for trade-off thinking. What does X mean in terms of Y? If I have only so much money, Z, how much of X or Y, or combinations of the two, should I buy? The reality is that despite the intellectual progression that would have budgets flow from strategy and military capabilities determined by objectives, they are seldom developed in that manner. Far more likely is that a budget will drive a strategy and that threats will determine which capabilities are deemed necessary. That being so the central questions of defense are "How much is enough?" and "To do what?" Boyd was always concerned about trade-offs and cost because cost has both immediate and long-term consequences.
>
> Beyond that, those considerations were the last that one should worry about. Boyd's trinity held people first, ideas second, and things third. Often the military has as its first priority the things, the high-tech weaponry. Ideas are second, and people, in that they are trained to be interchangeable parts, a tertiary consideration. That is not meant to seem as heartless as it sounds but merely to point out that we often seem to value the capabilities of our technology more than the people who use it…Boyd was convinced that one's mind was the best weapon, and hence, well-trained and well-educated people, who think well and quickly, were the most important asset, followed by ideas, in turn followed by the equipment they had at their disposal (p. 110).

While many of the claims on both sides of the debate appear to have face validity, it is beyond the scope of this summary to pursue them in more depth. What is important is to appreciate those aspects of the debate that touched on elements of what would become the Army's MANPRINT program. This is particularly true given the substantial leap forward the reform movement took in 1983, heralded by the appearance of reformer Chuck Spinney on the cover of *Time* magazine (Figure III-6). This, and subsequent events, helped catapult the military reformers' arguments into a wider national appeal for

military reform (Kross, 1985, pp. 161–180) at the same time that senior Army leaders were contemplating designs for the Army of Excellence.

Figure III-7. *Time* magazine cover from March 1983 featuring Franklin Spinney, a military analyst for the Pentagon who became famous in the early 1980s for his "Spinney Report" criticizing the perceived reckless pursuit of costly complex weapon systems by the Pentagon.

C. TECHNOLOGY AND ITS EFFECTS ON PERSONNEL

Martin Binkin (1986), in *Military Technology and Defense Manpower*, provides a good discussion of the effects of military technology on the occupational structure of the armed forces circa the 1980s. Written during the period without the benefit of hindsight, it captures much of the uncertainty and debate of the time that was likely prevalent in the minds of key actors in the MANPRINT story. Binkin asserts that the quantity and quality of personnel needed by the armed forces depends on several factors. These factors include the tasks that units are expected to perform (and hence their workload), how they are organized (i.e., combat-to-support ratio), and personnel policies (i.e., how people are assigned and utilized). The impact of technology comes into play when determining the number and qualifications of the personnel needed to operate, maintain, and support the

systems. Thus, when new systems and advanced technologies are fielded, the effects on the work force will depend, in large part, on the degree of system *complexity* (Binkin, 1986, pp. 43–44).

Although the term “complexity” is widely used to characterize military systems, those who use it seldom provide a clear definition of exactly what they mean. Some consider complexity as synonymous with job difficulty, gauging it by the time needed to learn a system’s tasks, as in training time. Alternatively, the quantity of documentation or job aids needed to support a system’s operation and maintenance is used as an indicator of complexity. Procurement costs also have been used as a measure of technical complexity. Finally, complexity has been defined as the number of components or parts that comprise a system (Binkin, 1986, pp. 44–45).

However it is defined, complexity generally carries a negative connotation because of the presumed inverse correlation between complexity and reliability, with the latter having important manpower implications. Although terminology often varies with the system being measured, reliability is usually defined in terms of the “mean time between failures” (MTBF). As a general rule, increasing the number of components in a system—and hence its complexity—results in a shorter MTBF and a greater need for maintenance. Another characteristic of systems that impacts manpower is maintainability, which is often measured in terms of the “mean time to repair” (MTTR). Although the MTTR is affected by factors other than complexity, it generally takes longer to diagnose failures in complex systems and it often takes longer to service components that are not readily accessible (Binkin, 1986, pp. 45–47)

The relationship between reliability and maintainability is captured in the concept of availability, which simply stated, is the proportion of time a system is in commission:

$$\text{availability} = \frac{\text{uptime}}{\text{uptime} + \text{downtime}}$$

Downtime includes both the time actually spent doing maintenance as well as the administrative and logistics delays incurred during the maintenance process, such as awaiting technical personnel, equipment, or spare parts. Downtime also encompasses maintenance actions that are unrelated to component failures, such as when scheduled or

preventive maintenance is performed to decrease the likelihood of system failures or retard wear. Thus, the number of people needed to maintain a particular system is a function of how often it fails (i.e., its reliability), how long repairs take (i.e., its maintainability), and the extent of scheduled maintenance. Collectively, these factors determine the maintenance workload, and consequently, the number of maintenance and support billets required (Binkin, 1986, pp. 47–49).

The balance of historical evidence, at least in the early 1980s, suggested that systems incorporating advanced technologies tended to be complex, and hence unreliable and hard to maintain with ensuing adverse affects on manpower. However, the question of whether past trends would be predicative of the future was a central one in the debates related to the military reform movement. Proponents for the new high technology weapon systems were convinced that historical trends would not bear out—emerging technologies would make military systems more reliable and easier to maintain. They embraced the idea of "transparent complexity" in which technology would enable user-friendly weapons and black-box, remove-and-replace maintenance concepts, thereby reducing both the quantity and quality of operators and maintainers needed for a given military capability. On the other hand, critics pointed to historical experience with high performance systems and cautioned that anticipated manpower savings from technological substitutions seldom materialized (Binkin, 1986, pp. 53–68).

What then was the outlook for the Army of Excellence? In summary, the substantial modernization that was well under way in the Army in the early 1980s was yielding weapon systems with embedded electronic and computer technologies that promised to vastly increase their capabilities—and if history was any guide, their complexity. Unlike the Air Force and the Navy, both of which had a long association with complex weapon systems, the Army had traditionally emphasized men over equipment (Binkin, 1986, p. 7). Consequently, the Army experienced significantly more turbulence as it converted from systems that were largely electromechanical to systems incorporating advanced integrated electronics. Whether the Army would actually realize the full performance designed into these systems was an open question at the time, the answer to which depended largely on the extent that Army personnel were up to the task

of operating and maintaining these new weapon systems (Binkin, 1986, p. 34). However, as the "bow wave" of force modernization swept through the Army in the early 1980s, the problems predicted by the reformers began to emerge:

> First, when a new complex system was put in the hands of soldiers, the overall system performance did not always meet the standards predicted during the engineering design. Second, the replacement of a fielded weapon system with a more technologically complex system was generating both a requirement for more highly skilled soldiers as well as a higher soldier to system ratio when counting the number of operators, maintainers, and support personnel (Blackwood & Riviello, 1994, pp. 2–3).

Figure III-8, which was adapted from Binkin (1986), highlights the growth in technical (i.e., combat support) jobs, which accounted for a quarter of all enlisted soldiers by 1985—a trend that was opposite of that intended by Army planners responsible for designing the Army of Excellence.

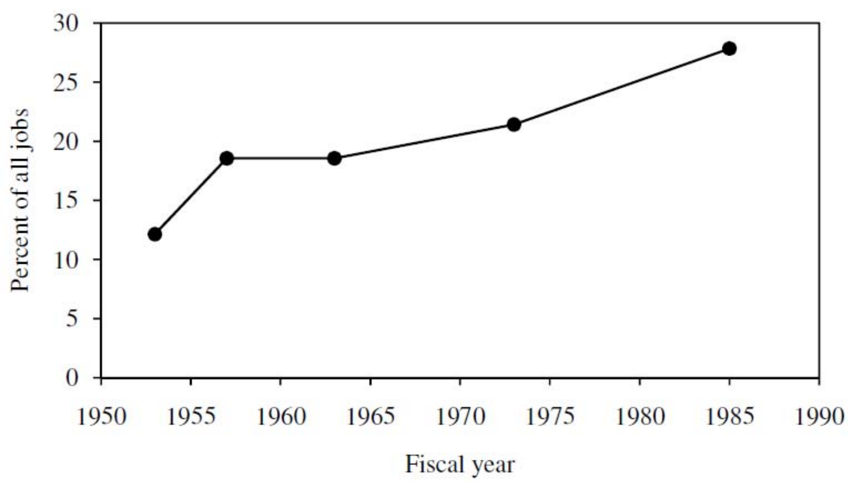


Figure III-8. Growth of technical jobs in the U.S. Army for selected fiscal years 1953–1985 (After Binkin, 1986).

As early as August 1980, Generals (retired) Walter Kerwin and George Blanchard highlighted concerns within the Army regarding force readiness, supportability, and sustainability. A World War II veteran, General Kerwin was instrumental as Deputy

Chief of Staff for Personnel in launching the Army's voluntary enlistment program in 1973 prior to serving as the Vice Chief of Staff from 1974-1978. Likewise, General Blanchard, a combat veteran of three wars, helped with the transition to an all-volunteer force as commander of U.S. Army, Europe. In an Army Materiel Systems Analysis Activity discussion paper prepared for Army Chief of Staff General Edward Meyer, Kerwin, Blanchard, and colleagues bluntly stated (Kerwin, Blanchard, Atzinger, & Topper, 1980):

> The U.S. Army has a major man/machine interface problem. There are not enough qualified people to perform the tasks required to effectively operate, support, and maintain current Army systems…The problem is severe and will continue to get worse. Increasing weapon complexity, the large number of new systems being developed, insufficient formal school training, a declining manpower pool, disproportionate numbers of CAT IIIB and CAT IV[6] personnel, recruiting and retention problems, and unit turbulence all will continue to strain the already overburdened personnel, training, and development communities (p. 2).

They went on to propose that the Army's leaders adopt a total system view that considered soldier performance and equipment reliability together as a system:

> The Army has made some progress in dealing with this problem. Many efforts are underway. However, these efforts, while representing steps in the right direction, are fragmented, based on reactions rather than vision, and, to a large extent, individually initiated. In our opinion, these efforts will fall short in coping with the extent of the problem in time to have an impact in the near term. Significant improvement will not occur quickly unless efforts are integrated, the personnel and doctrine people become more actively involved early in the materiel development process, and the Army addresses man/machine interface in its broadest sense and begins to think tactical system development in lieu of individual materiel development, individual people development and individual support development (p. 2).

Specific observations presented in the discussion paper included:

[6] The Armed Services Vocational Aptitude Battery (ASVAB) is a multiple choice test, administered by the United States Military Entrance Processing Command, used to determine qualification for enlistment in the United States armed forces. For enlistment purposes, scores are divided into several categories: Category I (93-100%), Category II (65-92%), Category IIIA (50-64%), Category IIIB (31-49%), Category IVA (21-30%), Category IVB (16-20%), Category IVC (10-15%), and Category V (0-9%). By Congressional mandate, no Category V recruits can be accessed into the military and no more than 20% of accessions can be Category IV.

- The Army's Life Cycle System Management Model must be disciplined concerning the manpower, personnel, and training (MPT) and logistics aspects of the process.
- Careful consideration of MPT impacts must precede any variation in strategy which skips a phase of development for the purpose of achieving an early Initial Operational Capability.
- MPT requirements must be better defined during concept evaluation.
- System development programs must recognize training constraints and employ sophisticated techniques to reduce training requirements.
- Human factors analysis and engineering must become a mandated part of system development early in the cycle.
- Program managers and U.S. Army TRADOC system managers must increase their emphasis on the MPT features of the Integrated Logistics Support process.
- The personnel community must become an active, rather than reactive, part of the acquisition process.
- The Army needs a central authority for integrating systems development, acquisition, and fielding.

Kerwin and Blanchard's paper was accompanied by other reports of contractors and Defense Department personnel failing to adequately consider manpower requirements and human factors during the weapon system acquisition process. For example, the U.S. General Accounting Office (GAO) issued its report, *Effectiveness of U.S. Forces Can Be Increased Through Improved Weapon System Design*, in January 1981, concluding that many military weapon systems could not be adequately operated, maintained, or supported because the Defense Department had not given sufficient attention to logistic support, human factors, and quality assurance during the design phase of the acquisition process. The GAO stated in their report to Congress:

> We believe there are three important ownership factors in the acquisition process which recent history suggests are among the most prominent detractors from the effectiveness of deployed systems - logistic support, *human reliability* [emphasis added], and quality assurance. Our selection of these ownership factors for analysis does not imply that others are

> unimportant. Rather, we suggest that there has been an imbalance of funding and attention given between the measurable characteristics of weapon system development (cost, schedule, and performance) and these other factors which significantly influence the eventual effectiveness of the system in the field (p. 4).

Specific observations in the report that related to human reliability included:

- There are indications that poor human reliability causes over 50 percent of all weapon system failures.
- The increasingly complicated nature of modern military systems, together with internal military personnel problems, suggests that human-induced errors, both in operations and maintenance, will likely increase unless more attention is paid to this problem during design and development.
- Weapon system designs have been dictating manpower requirements when what is needed is a continuous interface between the system designers and the manpower planners with manpower requirements influencing system design and vice versa.
- Human factor specifications, standards, and handbooks used in designing and developing systems almost exclusively deal with the human physical characteristics and design interface and do not adequately address other human limitations such as skill levels, proficiency, availability, environmental stress, and fatigue.
- There are no common methodologies and data sources for use by system designers in forecasting skill levels of future military personnel.
- There is insufficient emphasis on testing systems from a human reliability standpoint, particularly in the developmental stages of the acquisition process when design errors are easiest to correct.

While the military buildup of the early Reagan administration may have served to make these problems appear acute, they were really just a flaring of a more chronic pathology within the Defense Department. Going back to October 1967, the Office of the Director of Defense Research and Engineering, on the recommendation of the Assistant Secretary of Defense for Systems Analyses (ASD/SA), conducted a study of the

adequacy of consideration of manpower factors during the system acquisition process. One of the first problems that the study group had to tackle was one simply of terminology (Nucci, 1967):

> While the terms human engineering (HE) and human factors (HF) do relate to the man/machine interface, their application thus far in the system-design process has mainly dealt with man's inherent characteristics and capabilities. HE and HF do not, however, address such personnel factors as quantities, skill levels, proficiency, availability, and costs. Moreover, the HE/HF area does not adequately embrace training aspects (p. 4).

After defining manpower factors to broadly include human factors, human engineering, human resources, and training, the study group went on to examine six current system acquisition programs. Summarizing their major findings, the study group reported:

> Important gains are achievable when manpower considerations, systems analyses and tradeoff studies are part of the early development phases (concept formulation and contract definition, as well as advanced development). But these gains are inhibited unless (1) management attention ensures consideration of manpower requirements and the man/machine interface in the course of program approval; (2) manpower factors are an inherent part of systems engineering; (3) the program is conducted on the basis of total system effectiveness, all tradeoff studies, and the final design reflection of consideration of total life-cycle costs (LCC); and (4) manpower factors are reflected in contract requirements (p. 5).

The study group also took up the issue of coordination of efforts related to manpower factors:

> While the Military Services are sponsoring many activities in human factors research, manpower study, life-cycle costing, maintenance-data feedback, personnel and training, etc., there appears to be little intra-Service—much less inter-Service—coordination of this work, especially from the viewpoint of using results in systems development and analysis (p. 5).

The study group further concluded that the military policies and procedures concerning manpower factors were inadequate and that, while human engineering received the most emphasis because of contractual requirements, training and human resources received

little emphasis. Finally, the study group recommended designation of a Defense Department focal point for manpower factors, going so far as to propose responsibilities for such a focal point.

A decade later, in December 1978, the Logistics Manpower Institute concluded another study of manpower planning for new weapon systems, this time for the Assistant Secretary of Defense, Manpower, Reserve Affairs, and Logistics (ASD/MRA&L). In the report, the authors stated (Betaque, Kennelly, Nauta, & White, 1978):

> Until recently, there was a decided lack of specific guidance from the Office of the Secretary of Defense on manpower planning for new systems. That deficiency was corrected by a 17 August 1978 ASD(MRA&L) memorandum, "Manpower Analysis Requirements for System Acquisition." Consequently, DoD policy on manpower planning for new systems now appears adequate, but still there are serious shortcomings in the presentation and implementation of that policy (p. iii).

The study was complemented by seven case studies, two of which concerned Army systems. Significant findings from the latter included the following:

- Most estimates of manpower requirements made during acquisition programs were too low.
- There was greater uncertainty associated with maintenance manning than with any other element of new weapon system manpower requirements.
- Estimates of new system manpower requirements frequently reflected program goals rather than unbiased assessments of manpower needs.
- Manpower goals or constraints established for new systems addressed only the aggregate manning of the using unit, not total manpower or skill requirements.
- Controlling training requirements could be as important as constraining manning levels.

Binkin (1986) asserts that the implications of new technologies for future manpower requirements are best analyzed by looking at the effects that these technologies have on the operations and maintenance of military systems themselves and on the supporting infrastructure—that is, the total ownership considerations (pp. 39–40).

However, as evidenced by the aforementioned studies, such analyses generally received little emphasis by the military services (Binkin, 1986):

> Practically…such analyses have been hampered in the past by a lack of reliable data as well as by the low priority afforded by the services to manpower research in general and analysis of manpower requirements in particular. Since the end of conscription, research on military manpower has concentrated on supply issues. Numerous econometric recruitment and retention models have been developed, and there is no shortage of estimates of elasticities of supply with respect to the many variables thought to affect the recruitment and reenlistment decisions. The mountain of often widely conflicting estimates makes all the more conspicuous the paucity of studies relating to manpower requirements. Thus most of the attention given to changes in occupational mix has been directed toward the near term and usually as part of the annual budgeting process (p. 40).

The result was a reactive culture in which system designs drove manpower requirements instead of the proactive approach of using manpower considerations to derive system design criteria (Nucci, 1967).

In 1980, Chief of Staff of the Army, General Meyer, took the action of ordering an in-depth analysis of the impact on the increasing complexity of Army weapon systems on personnel, training, and logistic aspects of force modernization. The resulting *Soldier Machine Interface Requirements* (*"Complexity"*) *Study* was completed by TRADOC's Combined Arms Center in May 1982. After examining 20 new systems, the study group found that institutional training requirements associated with new systems were steadily increasing, both in terms of the number of courses and their length. These impacts were felt in both initial and sustainment training, especially for logistic-oriented skills. The study group also observed a general upward migration of aptitudinal skill requirements to satisfy the demands for operating and maintaining new, high-technology systems. This trend was most evident for maintenance and repair skills, where new systems were reflecting the transition of applied technology from mechanical to electronic.[7] An unintended consequence was the emergence of families of high-skill, low-density jobs

[7] As an extreme example, the initial entry aptitude score required for the operator/repairman of the AN/MSM-105, General Support Automatic Test Support System, was higher than that required to enter Officer Candidate School.

that then made for significant personnel management problems. Although new systems were generally easier to operate, they were significantly more difficult to repair and maintain. In summary, the growth in personnel overhead, both in terms of maintenance and training, resulting from force modernization was working to exacerbate—not improve— the tooth-to-tail ratio by increasing the numbers of personnel comprising the tail (Ostovich, Jordan, Fowler, & Hatlestad, 1982).

The "Complexity Study" also examined the acquisition process and found some explanations for the adverse impacts of complexity. Among their findings, the study group reported (Ostovich et al., 1982):

> Total system equipment requirements are not developed in a coordinated manner…the weapon, its support equipment, and its training devices were not developed in parallel…This tends to support the criticism that the Army does not know what resources are really required to field a complete system (pp. xxxi–xxxii).

They affirmed the need to "think system development [emphasis in original]" (p. xxxii) and accordingly state requirements as "man-machine-system minimum mission essential functions" (p. 376). Additionally, the report argued that the personnel community needed to become more proactive in addressing the problem:

> The combat effectiveness of new weapon systems will depend on the Army's ability to identify, attract, train, and retain soldiers who are both capable of operating and maintaining them. At the present, there is no single, coherent framework for adequately analyzing these needs. A consistent procedure for examining manpower and the necessary soldier qualifications to determine accessions, selection, training, assignment and reenlistment adequacy does not exist. The relationship of the soldier's ability to the requirements of new system is largely unknown. Relating these factors is vital to avoiding inappropriate manpower, personnel, and training decisions (p. 384).

Overall, the study group found that the Army could—and needed to—do a better job at integrating MPT considerations during system design and acquisition. They also recommended that Army regulations assign responsibility for providing MPT data to system developers and for coordinating, validating, and evaluating MPT considerations for major defense acquisition programs.

That same year, then Army Deputy Chief of Staff for Personnel, General Maxwell Thurman, directed the U.S. Army Research Institute to undertake the Reverse Engineering Project, a series of four case studies examining if and how human factors, manpower, personnel, and training (HMPT) issues were addressed during the weapon systems acquisition process. The term "reverse engineering" was used by General Thurman to imply the method of determining how products of the weapon system acquisition process came to be as they were. It was his premise that careful examination of several Army systems that had already been fielded would allow identification of critical events in the weapon system acquisition process where, if proper consideration were given to HMPT issues, the Army could increase the likelihood of fielding more operationally useful systems (Promisel, Hartel, Kaplan, Marcus, & Whitenburg, 1985).

The Reverse Engineering Project identified several recurring HMPT problems: human factors engineering was not addressed for some system components, doctrine and operational and organizational concepts were incomplete or ill-suited to the soldier, manpower levels were underestimated, skill and ability needs were undetermined or underestimated, training was untested, and training devices were unavailable. Numerous factors in the weapon system acquisition process were found to have directly produced these HMPT problems, but the project's authors concluded (Promisel et al., 1985):

> A core problem underlies the various factors related to HMPT problems…HMPT problems have their origin in inadequate or incomplete analysis of the proposed system during the concept stage. This leads both to incomplete specification of requirements and inappropriate assumptions regarding system features. HMPT design parameters and field test design become too narrowly defined. The incomplete field tests cannot identify comprehensively errors of commission and omission regarding HMPT. The end results are: (1) HMPT problems; and (2) uncertainty regarding the adequacy of system performance along with inconclusive evidence concerning the importance of HMPT problems (p. 35).

Specific recommendations from the Reverse Engineering Project included:

- Total system performance in the operational environment should be the focus of the weapon system acquisition process from initial analyses through testing and decision making.

- Past practices involving baseline comparison systems or standard procedures, particularly as they have been applied to personnel considerations, should not be adopted for new or successor systems without specific analysis of their applicability.
- Actions and documents in the acquisition process should not be approved until their comprehensiveness, including attention to HMPT, has been verified.
- There should be systematic monitoring of processes specified in requirements and planning documents (e.g., tradeoff studies).
- Acquisition process decisions with bearing on HMPT issues should reflect estimates of the cost-effectiveness and cost-benefits associated with the available options.
- Characteristics of the acquisition strategy (e.g., competition, accelerated development, etc.) and their impact on HMPT issues should be explicitly considered in acquisition process planning.
- Actions should be taken to reduce turnover and improve HMPT-related training of appropriate acquisition personnel.
- In selecting and monitoring contractors, the competence of their staff in terms of HMPT should be assured.
- Responsibility for development of the total system, including HMPT, should be centralized to the maximum extent possible.

While the study authors emphasized that the last recommendation was key, the more important message was the overall finding of the study regarding the feasibility of addressing HMPT considerations in the weapon system acquisition. In summary, the observed shortfalls in existing systems were not inevitable consequences of increased system complexity.

D. THE DEVELOPMENT OF MANPRINT

1. Progenitor Ideas (1930s–1950s)

Although a number of individuals in the period spanning the decades of the 1970s and 1980s offered ideas that laid the groundwork for the subsequent development of the

U.S. Army's MANPRINT program, it is evident that the basic ideas they were espousing were recognized in one form or another at least as far back as the late 1930s (W. O. Blackwood, personal communication March 3, 2010). As discussed by Guilmartin and Jacobowitz (1985) in their historical analysis of group interaction and the design of military technology, the integration of technology, organizational practices (i.e., military tactics), and social mechanisms that foster effective human action and interaction has been the primary historical determinant of military victory. They assert that these factors—technology, tactics, and human factors—interact with one another in a dynamic manner than is not explainable in simple additive terms. When the confluence of these factors is such that they are mutually reinforcing, the result is a force multiplying effect. And in the case where these factors interfere rather than reinforce, the end result is a negative or force dividing effect.

Mutually reinforcing confluences of these factors does not appear to be the historical norm; to the contrary, true force multiplier effects are decidedly uncommon (Guilmartin & Jacobowitz, 1985):

> In most countries, the weapons system design and procurement process makes no formal provision for considering the nature of the interaction between the social and psychological dynamics of the combatant group and the characteristics of the military technology it uses. Where human factors are considered—as they must be in a narrowly physiological sense—the tendency is to view weapon system operation as an athletic endeavor involving individuals or small groups working in isolation. The prevalent belief seems to be that if the operator or crew is well trained and capable, the system will realize its theoretical, quantifiable potential; if not, it will not (pp. 2–3).

As a historical illustration, Guilmartin and Jacobowitz use the case of the World War II era French SOMUA S35 medium tank (Figure III-9), which possessed perhaps the best blend of armor protection, armament, speed, and agility of any armored fighting vehicle in the world at the time. A comparison of the SOUMA tank to the German opposition in the winter of 1939-1940, using probability-of-kill as a figure of merit, would have judged the French tank the sure winner. However, an examination of the design of the French SOUMA tank and German tanks used in the Battle of France highlight important differences in the quality of crew cohesion:

> Where German tanks almost invariably had the crew grouped together in a large and relatively spacious central compartment, French tank designers tended to isolate the individual members of the tank crew. German tanks…all had three-man turrets; French tanks had one-man turrets. German tank designers favored a side-by-side seating arrangement for the driver and assistant driver (who doubled as bow gunner/radio operator). With this setup, the two men could see and communicate with each other across the hull. By contrast, the crew members of French tanks (there was rarely an assistant driver) tended to be arranged in tandem and separated by machinery. […] Above and beyond the relative isolation of individual French tank crew members notwithstanding, tactical coordination within units was rendered tenuous at the outset by the physical arrangements faced by tank and unit commanders. So was cohesion. Fighting from a one man turret, the French tank commander was responsible for loading, aiming, and firing the primary armament…in addition to commanding their tanks and leading their units (pp. 49–50).

In the end, the negative confluence of high technology, tactics, and human factors—the latter in terms of the all-important element of primary military group cohesion—led to the quick defeat of French armored forces.

Figure III-9. French SOMUA S35s captured by Germany in 1940.

The lesson in this case exemplifies Guilmartin and Jacobowitz's premise that "a weapon…can, through the characteristics of the employment tactics, enmesh itself with the social and psychological nature of the soldiers who use it to reinforce the primary bonds which are the basis of the fighter's ability to withstand the stresses of combat and the horrors of battle" (p. 7). Whether one wishes to accord such a statement the status of an enduring principle, the historical record suggests, at a minimum, that the general notion was grasped and applied to weapon system design by the German *Wehrmacht* in the 1930s. It is also worth noting that the major elements under consideration could be reasonably argued as foreshadowing the holistic focus on human factors engineering, personnel survivability, and habitability considerations that was to become a major thrust of the U.S. Defense Department's future HSI program.

Nevertheless, the emergence of a discipline specifically focused on human factors can best be regarded as an outgrowth of aviation in World War II. While there are many subplots within the larger story of the history of human factors, one in particular—John Flanagan's development of the critical incident technique—stands out because of its continued evolution and application for the study and design of human performance interventions across a wide range of situations. Flanagan's critical incident technique is best regarded as a product of the many studies conducted in the U.S. Army Air Forces Aviation Psychology Program, which was established in the summer of 1941 to develop procedures for the selection and classification of aircrew. The technique itself is essentially a set of procedures for collecting direct observations of human behavior in such a way as to facilitate their potential usefulness in solving practical problems and developing broad psychological principles (Flanagan, 1954, pp. 327–329).

Given the success of the technique during World War II in analyzing a range of activities to include combat leadership and spatial disorientation in pilots, it was further developed during the postwar period, primarily by the American Institute for Research and the University of Pittsburgh. During the 1950s, applications of the critical incident technique were expanded to include studies in the following areas: measures of typical performance (i.e., criteria), measures of proficiency (i.e., standard samples), training, selection and classification, job design, operating procedures, equipment design,

motivation and leadership, and counseling and psychotherapy (Flanagan, 1954, pp 330–334, 347–355). In so doing, Flanagan and colleagues established the paradigm of subjecting human dependent variables to historical analysis for the purpose of generating hypotheses to apply to the present. Moreover, in the process, they developed functional descriptions of a broad range of human activities within complex systems based on a variety of perspectives, many of which would align with future MANPRINT domains.

2. Failed Initial Attempts (1967–1983)

Dr. John D. Weisz was a prominent figure in the Army human factors community during the period from the 1950s through the 1980s.[8] An infantryman in World War II, Weisz earned a doctoral degree in experimental psychology after the war, marking him as a member of the second generation of human factors specialists. Weisz spent his career working in the Army's Human Engineering Laboratory (HEL), which itself was only established in 1952 to assist in the development of engineering designs so that soldiers could use their equipment "in the best possible way." During the 1962 Army reorganization, HEL became a corporate laboratory within the Army Material Command and was charged with coordinating all the human factors engineering initiatives within the Army (Army Research Laboratory, 2003, p. 11). Consequently, as the Army Material Command started to emphasize systems analysis[9] in the 1960s, HEL began to actively consider the human factors contributions to such analyses. In the words of then HEL director Weisz (1967), "Since it appears that system analysis will become a standard technique in research and development, especially whenever there is a choice between two different proposed system concepts, or various alternative system mixes to meet a particular threat, it is now appropriate to determine what role human factors researchers

[8] Except where otherwise noted, section D is substantially based on a compilation of a series of interviews carried out from January–March 2010. These interviews, conducted by the author, included the following people: William Blackwood, Kenneth Boff, Paul Chatelier, and Joyce Shields.

[9] As discussed by Hughes (1998), in the 1950s and 1960s, the definitions of the terms systems engineering, operations research, and systems analysis frequently overlapped. Usually, those speaking of systems engineering had in mind either the management of the design and development of systems with purely technical components or sociotechnical systems with both technical and organizational components. Proponents of operations research referred usually to quantitative techniques used to analyze deployed military and industrial systems. Those that were expert in systems analysis compared, contrasted, and evaluated proposed projects, especially those that would create weapon systems (p. 142).

should play in making the analysis" (p. 1). Given Weisz also mentions recent proposals to use systems analysis to aid logisticians, he was probably responding, at least in part, to the findings and recommendations of the Department of the Army Board of Inquiry on the Army Logistic System, or the Brown Board, so named for the board's chairman, Lieutenant General Frederic Brown.

The Brown Board (Department of the Army, 1967) was established by the Army in September 1965 for the purpose of analyzing the Army logistics system to determine what changes and modifications were needed to make it more responsive to materiel readiness requirements. The need for such an analysis was driven, in part, by the failures of previous logistics reorganizations to improve unit-level readiness and the perceived loss of "the personnel-training-doctrine-hardware" (p. I-6) integration function that was historically accomplished by the Army's Technical Services, the latter having been largely eliminated during the Project 80 reorganization.[10] Part of the Board's study approach was to set up a systems analysis and operations research group to describe and analyze Army logistics as a system. Not surprisingly, a major concept advanced by the Board was the need for the Army to use systems analytic and operations research tools as part of its internal management techniques:

> DOD employs the management tools of systems analysis, operations research, statistical analysis, cost effectiveness, and other technical disciplines. The Army must also use these tools if it is to respond articulately to DOD. More than this, DOD uses these tools because they

[10] To achieve increased economies in the common or cross service support areas, a series of Defense Department studies were directed to explore areas that might result in more cost-effective support operations. One of these, Project 80, which established the Hoelscher committee, was directed to review the Army organization, with particular emphasis on the logistics structure. Some of the principal findings of the Hoelscher committee centered on the Army's seven Technical Services, each of which operated a distinctive, vertically integrated logistics system keyed to the nature of commodities provided and services performed. As a group, these services performed the Army's wholesale logistics functions, included research and development, procurement and production, inventory management, storage and distribution, maintenance and disposal, technical and professional services, and training and development. The Hoelscher committee proposed a consolidation of materiel functions and transportation services under a single Department of the Army operational command that was later named the Army Materiel Command (AMC) [In the actual implementation of the Hoelscher committee proposals, transportation services were not placed under AMC]. As a result of the reorganization, the personnel, training, military occupational specialty proponency, and doctrine functions formerly performed by the Technical Services were transferred to the Office of Personnel Operations, the Continental Army Command (CONARC), and the Combat Developments Command (CDC) (Department of the Army, 1967, pp. I-3 – I-5).

> have been proven to be effective techniques for quantitatively structuring decision making and resource allocation in complex operations (p. III-2).

Additionally, the Board suggested the need for the Army General Staff to develop a coordinated set of principles and policies for logistics using a "systems approach:"

> A systems analysis approach to the management of Army logistics must be applied to establish the basis for a coordinated and compatible set of principles and policies. All aspects of Army logistics should be described using the tools of systems analysis…A model, preferably a mathematical one, should be developed of the total logistics system (p. III-4).

In summary, the Board's approach and recommendations mirrored the emphasis on systems analysis that was pervading the Defense Department under Secretary of Defense Robert McNamara, himself an ardent proponent of the approach.

Beyond the issue of methodologies, the Board (Department of the Army, 1966) found that "one of the most damaging weaknesses in the acquisition-of-materiel cycle [was] the failure to plan for and provide the nonmaterial aspects of equipment" (p. III-7). A key element of the Board's solution to this problem was a systems analysis approach to the management of all interrelated materiel and nonmateriel activities:

> The focus of the systems analysis approach…was on the management of the process necessary to logically consider and evaluate each of the military, technical, and economic variables involved in the total system design which will achieve stated requirements. The creation of a balanced system design demands that each major design decision be based on an appropriate consideration of systems variables. These include equipment, facilities, personnel requirements, procedural data, training, testing, follow-on logistics, support, and intersystem and intrasystem interfaces (p. VI-9).

Additionally, the Board found that "there [was] no cohesive human factors engineering program in the Army materiel development programs" (p. VI-16), nor did the regulations "provide adequate guidance to establish a basis for [human factors engineering] requirements in system development (i.e., system analysis of human performance requirements)" (p. VI-17). Also, the materiel acquisition process was unable to assure the availability of qualified personnel at the time materiel was deployed because of inadequacies in the processes by which personnel, training, and manpower requirements

were developed. Thus, the Board recommended establishment of a comprehensive human factors engineering program to include issuance of a related Army Regulation. The Board also recommended "[c]onsolidation of personnel, training, and organizational functions under a single command…[to facilitate the] timely provision of qualified personnel" (p. IX–28).

Coming on the heels of the Army's Brown Board study was a Defense Department study (Nucci, 1967), summarized earlier, highlighting the need to better integrate manpower factors (i.e., manpower, skill levels, proficiency, availability, rotation rates, costs, etc.) into materiel acquisition programs. It is therefore not surprising that Weisz (1967, 1968) perceived that it was the *de facto* policy of the Defense Department and the Department of the Army to take manpower factors into consideration in system analysis studies. Weisz (1967) intuitively understood the simple fact that, "No weapon system analysis can be considered valid unless it includes the contribution of a very important component, namely 'man,' as part of the system" (p. 1). He also realized that the behavioral sciences and human factors engineering fields, which historically utilized experimental methodologies and statistical techniques drawn from experimental psychology, needed to include operations research techniques within their list of tools for performing their portion of weapon system analyses. Weisz saw the field of human factors making contributions to the following four areas of a weapon system analysis: manpower requirements,[11] training requirements, performance requirements,[12] and system design. Finally, Weisz asserted that human factors should not be treated separately from other areas embraced in the system analytic thinking process. In his words, "Since man is an integral part of the total system, his contributions must be included each and every time that such areas as <u>system performance</u>, <u>system effectiveness</u>, <u>system dependability</u>, <u>system reliability</u>, <u>system capability</u>, and <u>cost effectiveness</u> are considered [emphasis in

[11] Weisz (1967) considered manpower the "critical commodity" and thus it needed to be considered for all proposed system concepts in terms of the number of personnel to be used and the skill levels of the personnel required to operate and maintain the system (p. 2).

[12] Weisz (1967) asserted that the contribution of man to system performance must always be considered in the particular environment in which the system will be utilized. Thus, environmental factors were incorporated in system analyses through their effect on man doing his tasks as part of total system operation (p. 2).

original]" (Weisz, 1967, p. 3). To that end, Weisz argued that man's contributions to these areas could be determined and expressed quantitatively as accurately at most of the other determinants (Weisz, 1967, p. 3).

Weisz published two HEL reports (*System Analysis: Human Factors Research and Development Contributions* [Technical Note 6-67, July 1967] and *System Analysis: Manpower Resources/System Design Integration Methodology* [Technical Note 9-68, August 1968]) establishing a general framework around which manpower factors could be effectively included and appropriately weighted in the materiel acquisition process (Weisz, 1989, p. 2). Moreover, an implementing guide (*Manpower Resources Integration for Army Materiel Development* [HEL Guide 1-69, January 1969]) was also published for the purpose of explaining how to integrate HEL's human factors engineering program within the Army's life-cycle management model. Concurrently, the first version of Army Regulation (AR) 602-1, titled *Man-Materiel Systems—Human Factors Engineering Program*, was published in March 1968 with the aid of HEL, laying out a Department of the Army-wide human factors engineering program. Noteworthy in this regulation was the redefinition of human factors engineering as "a comprehensive technical effort to integrate all manpower characteristics (personnel skills, training implications, behavioral reactions, human performance, anthropometric data, and biomedical factors) into all Army systems" (p. 1) to ensure operational effectiveness, safety, and freedom from health hazards. This definition was written by Weisz with the help of Jacob Barber, who worked in the Office of the Deputy Chief of Staff for Personnel, Research and Studies Office. Their intent was to broaden the scope of the definition of human factors engineering in use at the time (HEL Guide 1-69, January, 1969):

> The term "HFE" in the sense in which it is used in AR 602-1…encompasses all of the "human" factors with which materiel developers must be concerned. Although separate Army agencies exist to handle the selection, classification, and training of personnel who will ultimately operate and maintain new equipment, it is important that the aspects of manpower resources administered by those agencies be considered during materiel development (p. 1).

The various HEL publications were put forth in concert with AR 602-1 in support of what was really a pioneering endeavor to integrate all soldier-related attributes or characteristics into the materiel acquisition process (Weisz, 1989, p. 2).

Weisz's persuasive advocacy for his innovative ideas to integrate human factors into weapon systems analyses stimulated other organizations, including the Army's materiel development community, to at least imitate his approach. For instance, the U.S. Army Materiel Development and Readiness Command (DARCOM) reproduced large portions of Weisz's report, *System Analysis: Manpower Resources/System Design Integration Methodology*, verbatim in DARCOM Pamphlet 706-102, *Engineering Design Handbook, Army Weapon System Analysis, Part 2* (1979), under the heading, "Introduction to Human Factors and Weapons Systems Analysis Interface Problems." The handbook provided a rather eloquent introduction to the role of the weapon systems analyst in addressing human factors:

> Man has often been looked upon as one of the primary components of a complete weapon system, and his interface with the weapon or weapon system he employs may likely strike the balance between success and defeat in battle. It is a very natural approach for the analyst to identify each anticipated source of variation in expected performance of the weapon system, to estimate the relative sizes of the components of variation, and to find their effect on predicted overall system performance. Should it turn out that the man is contributing too large a component of variation toward expected system performance, then further training of the soldier, or operator, may be indicated, or perhaps the improved design of weapons will be made mandatory. Otherwise, it becomes important to estimate the size of natural human variations which may be involved in operation of a weapon system, and to take such amounts of variability into account in effectiveness studies (p. 33-3).

It then went on to identify three areas that should be of primary concern for the weapons systems analyst: human engineering, human performance, and human reliability. This was subsequently followed by examples of several evaluations for the purpose of introducing the analyst to pertinent aspects of these three areas. Even by today's standards, the handbook provided a highly memorable and salient introduction to the topic of human factors in weapon systems analysis.

In the end, however, Weisz and his HEL colleagues' efforts to develop what could essentially be characterized as the MANPRINT program of the mid 1980s amounted to a bottom-up campaign for reform that failed to garner sufficient support from the Army's senior leaders. As described by Weisz (1986):

> Though we certainly tried, we could not fully implement AR 602-1 in 1968 nor [sic] later on because of organization and policy problems. There was no mechanism to integrate the various aspects of the [AR 602-1 human factors engineering] definition across the responsible commands and agencies, nor sufficient policy and guidance to enforce such integration (p. 3).

Nevertheless, their efforts were not without future effect. In the June 1976 update to AR 602-1, Figure A-1 of the Appendix (reproduced as Figure III-10 below) was added to further detail the personnel considerations in system effectiveness. Figure A-1 was clearly the progenitor of a similar figure published 15 years later in DoD Instruction 5000.2 (1991) under the title "Human Systems Integration" (p. 7-B-2). Additionally, the HEL later prepared an updated version of HEL Guide 1-69, titled *Human Factors Engineering in Research, Development, and Acquisition* (October 1980), at the request of Generals Kerwin and Blanchard (authors of *Man/Machine Interface—A Growing Crisis*) after they were shown HEL Guide 1-69, which by then had become outdated because of changes in the materiel acquisition process (Weisz, 1989, p. 2). Moreover, Mr. Delbert L. Spurlock, who was Assistant Secretary of the Army for Manpower and Reserve Affairs from 1984–1989, was oft to quote Weisz's earlier publications (1967 and 1968) to the House Armed Services Committee in seeking support for the Army's then nascent MANPRINT effort (Spurlock as quoted in Patrick, 1988, p. 4; Weisz, 1986, p. 4).

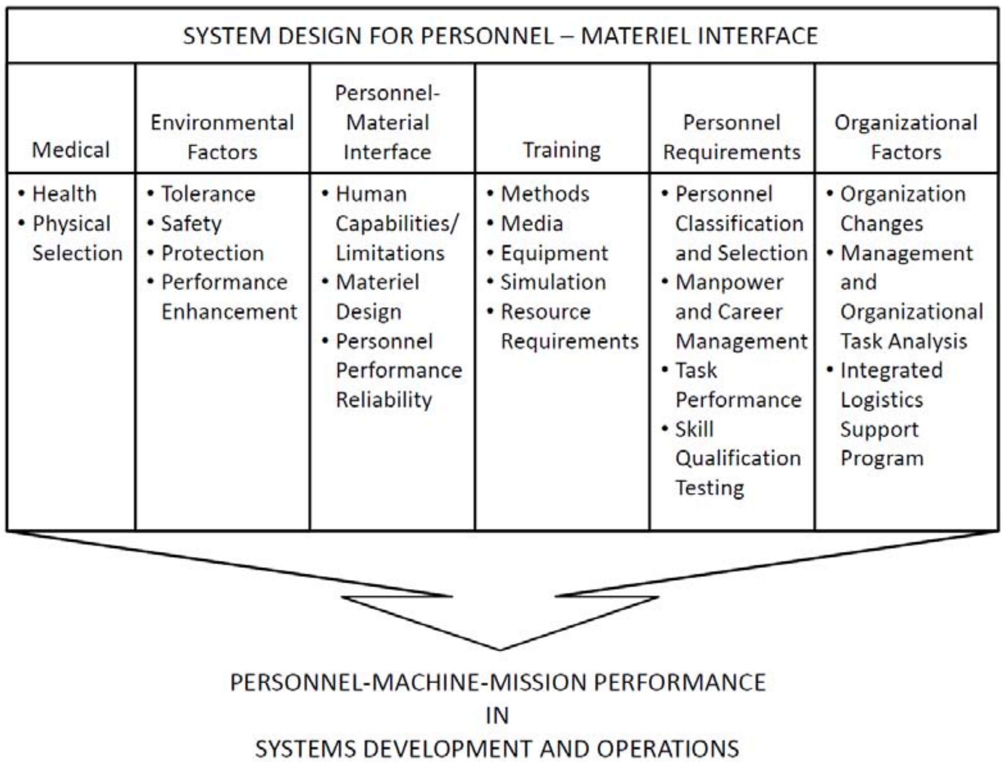


Figure III-10. Personnel considerations in system effectiveness (From Department of the Army, 1976).

3. A Resurgence of Interest (1983–)

Neil Sheehan, in his book, *A Fiery Peace in a Cold War: Bernard Schriever and the Ultimate Weapon* (2009), alludes to an old Marine Corps adage that "luck occurs when preparation and opportunity coincide" (p. 225). So it was to be with one U.S. Army Major William O. Blackwood. During the summer of 1983, with diplomacy between the U.S. and U.S.S.R. at the lowest point since Stalin's days and with the Army struggling to find a workable force structure, Blackwood reported for duty in the Research and Studies Office within the Office of the Deputy Chief of Staff for Personnel (DCSPER) in Washington, D.C. A career Army infantry officer with two tours in Vietnam, Blackwood earned a doctoral degree in education from the University of Florida in 1977 and was a scholar of organizational change. His assignments prior to the Office of the DCSPER had been with mechanized infantry divisions in Europe, and he

brought with him a close appreciation of the challenges soldiers faced with the new weapon systems of the time. As an infantry company commander responsible for testing new weapons, he gained a deeper understanding of the importance of man-machine interaction as a determinant of weapon system performance. Later, as a battalion operations officer, Blackwood witnessed firsthand the problems in training soldiers to high levels of proficiency with the Dragon Missile system and the lack of correlation between soldier proficiency on the training device and soldier performance with the weapon. Whether by chance, fate, or some purposeful design, Blackwood's subsequent assignment to Office of the DCSPER would turn out to be a proverbial case of the right person in the right place at the right time. Over the ensuing six years, he would apply his expert knowledge of organizational change to successfully effect how the military addressed personnel issues within materiel acquisition programs. Although having no formal training in human factors, Blackwood became the *de facto* architect and author of the Army's MANPRINT program during the period from 1983–1987. Later, during his follow-on assignment (1987–1989) to the Strategic Planning Office within the Office of the Undersecretary of Defense (Acquisition) he would serve as the catalyst for the introduction of HSI within the Defense Department at large (W. O. Blackwood, personal communication March 3, 2010, e-mail communication March 22, 2010).

While Major Blackwood was prepared by way of his education and experience, the coincidence with opportunity occurred when the normal rotation of the Army's uniformed senior leadership fortuitously brought two particular general officers to key positions from which they could sponsor organizational change. The first of these officers was General "Mad Max" Maxwell Thurman, a reputed workaholic and master organizer, who served as commander of the Army's Recruiting Command from 1979 – 1981 and then as DCSPER from 1981–1983. He is credited during this period with reversing the downward slide in Army recruit quality as well as for developing the Army's well-known "Be all that you can be" recruiting campaign. In 1983, Thurman became the Vice Chief of Staff of the Army, a position that put him on the Army Systems Acquisition Review Council, the latter having responsibility for reviewing major

acquisition programs at key milestones. Thus, there would now be a personnel-minded person sitting in judgment over major weapon system acquisition programs.

The other significant change in assignment involved Lieutenant General Robert Elton, who moved from being the commanding general of the 9th Infantry Division to take the DCSPER position vacated by Thurman. As discussed earlier, the 9th Infantry Division functioned as the High Technology Test Bed for development of a light infantry division design. Consequently, Elton had been dual–designated as the division commander and the High Technology Test Bed test director, which necessitated that he work in close coordination with both TRADOC and the Army Materiel and Development Readiness Command to marry up emerging technologies and organizational designs. Working for Chief of Staff of the Army, General Meyer, Elton had been granted relatively free reign to develop high technology light division designs and ideas independently of existing force planning efforts. Consequently, the new DCSPER brought to the position an innovative mindset, experience with both force design and materiel systems development, and a firsthand appreciation of the challenges the Army faced in launching the 10,000-man light infantry division project (W. O. Blackwood, personal communication March 3, 2010).

Thus the Army's personnel community had scored a trifecta in the summer of 1983. During this period, the new Chief of Staff of the Army, General Wickham, strongly pushed his ideas for the Army of Excellence (AOE), for which he had begun laying the groundwork in the spring of 1983 as Vice Chief of Staff of the Army. While it was Wickham who inaugurated the AOE design and gave it push and drive throughout, it was General Thurman, as the new Vice Chief of Staff of the Army, who was responsible for integrating the efforts of the Army Staff, TRADOC, the Army Materiel Command, and other major Army commands. The Summer 1983 Army Commanders' Conference synchronized the Army's leadership in terms of thinking about the resources needed to achieve the AOE design objectives. Above all, the light infantry division was the linchpin of the 1983 AOE design effort, and it was only going to be realized within the Army's apparently immutable end-strength ceiling by economizing on the use of manpower, especially in combat service support (Romjue, 1993, pp. 31–35). At the

same time, there were many examples of new weapons that did not perform well or increased rather than saved manpower spaces. Since the personnel community managed the Army's human resources, it was clear to General Thurman that they needed to be a full partner in this process. Consequently, Lt. General Elton tasked his Research and Studies Office in the summer of 1983 to develop a plan for giving the personnel community a "sense of place and purpose" in the weapon system acquisition process with the goal of improving manpower and personnel utilization within the Army (Blackwood & Riviello, 1994, pp. 3–4, 9)

It was readily apparent in 1983 from previous studies and analyses that the personnel community was not participating effectively in the Army's weapon system acquisition process despite the fact that they had the authority to do so since the 1970s. Several factors contributed to this situation: 1) a general lack of knowledge about the acquisition process by those in the personnel community, 2) inadequate analytical tools and techniques, 3) deficiencies in structured processes and accountability, and 4) the bureaucratic entrenchment of the acquisition community (Blackwood & Riviello, 1994, p. 6). In terms of the latter, there was particularly stiff resistance from the integrated logistics support (ILS) community who saw themselves as the focal point and management integrator for MPT actions (Blackwood & Riviello, 1994, p. 4). However, within the Office of the DCSPER, it was clear that a potential window for change had opened (Blackwood & Riviello, 1994, p. 7). The Army's senior leadership was ready for organizational change that improved integration of soldier considerations within the weapon system acquisition process—a prerequisite to achieving General Wickham's AOE design given the resource constrained environment (Stanley, 1985, p. 7). Additionally, people having both the appropriate *Weltanschauung* (i.e., worldview) and authority were in positions within the Army's senior leadership such that they could effectively sponsor significant organizational change (Blackwood & Riviello, 1994, p. 4).

Those in the Office of the DCSPER firmly believed that the answer to the question of who should be the integrator was not the ILS community. As they saw it, the personnel community was the one left to manage the long-term consequences of personnel-related decisions made in the weapon system acquisition process, so they

needed to be, at a minimum, a full partner in the acquisition process. The real question then was how to go about orchestrating the needed organizational change. Planning the change effort was carried out at the Office of the DCSPER by a small group of just three officers under Lt. General Elton, headed by Major Blackwood, and residing in the DCSPER's Office of Research and Studies. Their first step was to achieve consensus on an approach, which directly led to the decision by the Office of the DCSPER to sponsor an Army Science Board summer study (W. O. Blackwood, personal communication March 3, 2010).

There were several considerations that underlay the decision to orchestrate the change effort beginning with an Army Science Board study. The Army Science Board was seen as an essential means for validating the need for change and nominating an acceptable approach to solving the problem. Since the science board members included retired general officers and respected members of industry and academia, its findings and recommendations would have significant visibility among the Army's senior leadership as well as industry. Additionally, its members would likely be viewed as knowledgeable but impartial by the bureaucratic entrenchment, thereby helping to overcome anticipated resistance, particularly in the weapon system acquisition community. Perhaps most important, however, was the belief that the study recommendations would, if properly handled, eventuate in the Office of the DCSPER being empowered by the Chief of Staff of the Army to lead the Army Staff efforts to remedy the situation (Blackwood & Riviello, 1994, pp. 6, 10).

There was precedent for the Army Science Board to take up this issue as a logical extension of the Army Science Board Summer Studies 1981, *Equipping the Army 1990 – 2000*; 1982, *Science and Engineering Personnel*; and 1983, *Future Development Goals*, all of which discussed, in one form or another, the need to better integrate the soldier into the materiel acquisition process. Accordingly, the Army Science Board's 1984 Summer Study, *Leading and Manning Army 21*, took up the theme of improving readiness through better integration of personnel with the total Army system. One of the major study areas was the "soldier-machine interface," a term that was synonymous with human factors, manpower, personnel, and training (HMPT):

> The soldier-machine interface stretches across the boundaries of several technical disciplines and is now used to describe any number of often disparate approaches to systems design and analysis, logistics support analysis, and manpower planning…this proliferation of meanings has caused some to question the usefulness of the soldier-machine interface concept and others to relegate it to the imprecision of slang. The term, however, does meaningfully describe a specific methodology for improving systems design in the defense system development and acquisition process. This strategy fully integrates an emerging system's hardware, software, human, and other support subsystems in order to achieve specified mission capabilities…Hence, soldier-machine interface is a robust, yet precise concept, always useful and often required in order to optimize defense systems' design and ultimately their performance in the field (p. 121).

The study board found that institutional interest in the soldier-machine interface concept was driven in large part by the coincidence of several macroergonomic trends:

> The Department of Defense (DoD) has grown increasing reliant on advanced technology to counter the threat from a numerically superior potential adversary. By almost any measure, the density of high technology is up in the Armed Forces. […] In effect, the advanced operational capability of high technology systems has been purchased, at least in part, with greater demands for human resource. Yet the absolute size of the American workforce is shrinking. […] There has also been an alarming dip in the quality or capability of this smaller pool. […] This coincidence of a smaller, less capable workforce and burgeoning high technology in defense systems is already creating severe problems in military human resources and systems acquisition management. It is also impacting negatively on the combat readiness of the Armed Forces (pp. 121–122).

From a sociotechnical systems perspective, the science board's comments were clearly indicative of the Army having failed to jointly optimize its personnel and technological systems given the changes to its environment. Moreover, the comments provided a compelling *refutation* of Cold War technological determinism.

Having roundly validated the problem as seen from the perspective of the Office of the DCSPER, the science board then provided their suggested approach for solving the problem:

> In DoD, at least, the human-machine mismatch problem is as much a function of the way new defense systems are designed and developed as it

> is a product of shifts in the American population. Consequently, the solution requires a broad range of initiatives, involving both human resources and acquisition management. In order to increase total system effectiveness, DoD needs to simplify system operation and maintenance, and to reduce manpower requirements, training time, and cost. […] In effect, then, soldier-machine interface is a strategy for total system development [emphasis in original] (pp 123).

The science board went on to describe the total system development strategy:

> In the short term, this strategy will involve ad hoc actions in the system design process to ensure that emerging equipment is both affordable and supportable from a human resource perspective. (Said another way, the Services must take steps to ensure that they can efficiently access, train, and retain adequate numbers of personnel to operate and maintain new systems effectively.) These actions include training developments, personnel management, systems engineering, human factors engineering, and medical science.
>
> The initiatives are ad hoc in that they represent corrective and essentially independent efforts to redress immediate problems at the soldier-machine interface…In the longer term, the soldier-machine interface must extend beyond these useful but disconnected efforts to an alternative concept of system design. This concept is best characterized as total system development. […] This philosophy begins with a different concept of the final product of design—the system. With total system development, that product is a means to an end. It is software, hardware, human beings with logistics support, all of which must be creatively brought together to provide a desired mission capability. The emphasis, then, is on achieving field performance, rather than improving equipment, because only the former is genuine defense capability.
>
> The process of creating a comprehensive system that provides a desired capability requires a working integration of all technical disciplines involved with the system during its life cycle. […] The total system development process extends responsibility for both system design and field performance to all of these disciplines. […] Total system development will also require changes in DoD's investment philosophy. […] Ultimately, the traditionally horizontal approach to systems development (i.e., from concept definition, through concept demonstration and validation to full scale engineering development) will shift to a more integrated and vertical strategy. […]
>
> Fundamentally, total systems development will necessitate a new way of thinking about systems, a philosophy which focuses on the system's purpose rather than on the specifications, standards, goals and objectives,

> however detailed, of its constituent components. A good fit at the soldier-machine interface pushes technology and human and other support resources to their collective limits in pursuit of mission capability. Hitachi has called this concept "humanication"; others have described it as "equipping the man" [emphasis in original] (pp. 124–126).

Providing some further specificity to the total system development strategy, the science board identified several key functional areas—what would later become the MANPRINT domains[13]—that must necessarily be considered holistically in the system design process:

> The design objectives of a total system are achieved by integrating into the engineering design process the elements of human factors engineering, manpower, personnel, training and training equipment (HMPT). Also inherent in HMPT are many biomedical aspects and health hazard

---

[13] The original MANPRINT domains (Booher, 1990):

- Manpower – The number of human resources, both men and women, military and civilian, required and available to operate and maintain Army systems.
- Personnel – The aptitudes, experience, and other human characteristics necessary to achieve optimal system performance.
- Training – The requisite knowledge, skills, and abilities needed by the available personnel to operate and maintain systems under operational conditions.
- Human factors engineering – The comprehensive integration of human characteristics into system definition, design, development, and evaluation to optimize the performance of human-machine combinations.
- System safety – The inherent ability of the system to be used, operated, and maintained without accidental injury to personnel.
- Health hazards – Inherent condition in the operation or use of a system (e.g., shock, recoil, vibration, toxic fumes, radiation, noise) that can cause death, injury, illness, disability, or reduce job performance of personnel.

The Army Science Board 1984 Summer Study only addressed six domains. In the wake of Operation Desert Storm (1990–1991), an important lesson learned was that fratricide had to be reduced. Chief of Staff of the Army, General Gordon R. Sullivan, affirmed that the Army would not tolerate casualties that could be prevented by proper research, development, and acquisition, thereby focusing attention on the issue. Many believed that soldier survivability was a subset of system survivability, the implication being that if the system survives, so does the soldier. However, this is not always the case. In 1992, the DCSPER, Lt. General Thomas P. Carney, proposed resolving the issue by including "soldier survivability" as a seventh domain in the Army's MANPRINT program. In 1994, the U.S. Army Research Laboratory was given responsibility for soldier survivability and it was officially added as the seventh domain of MANPRINT. Thus, the above list was appended with the following additional domain description:

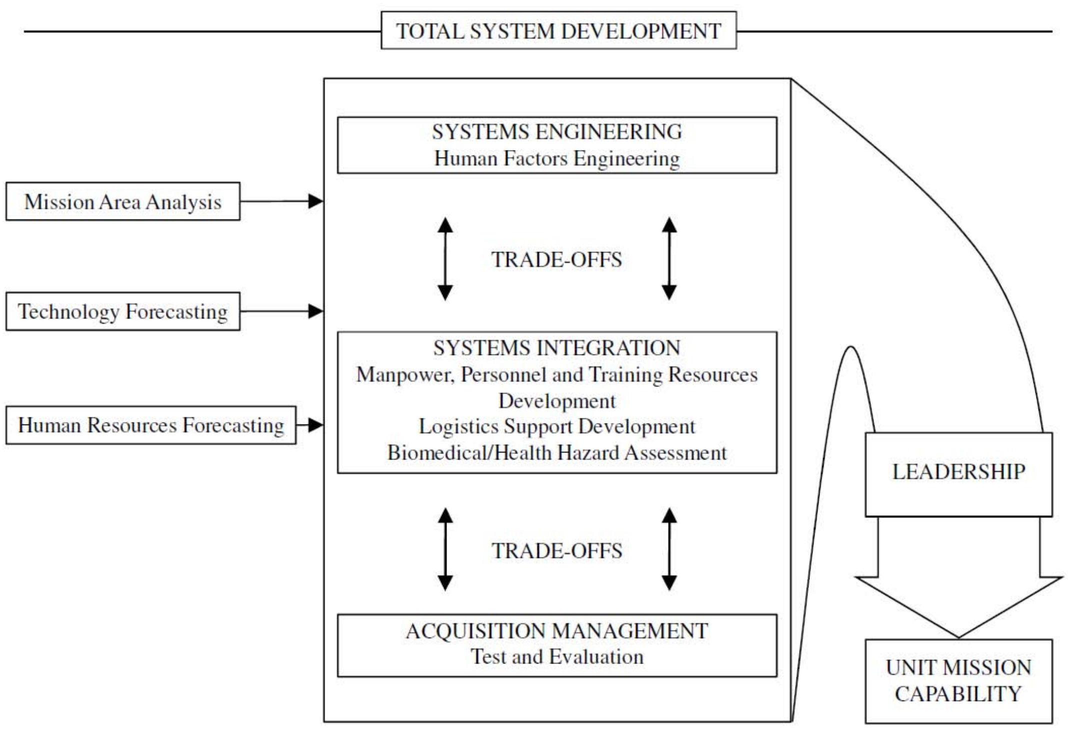
- Soldier survivability – The characteristics of a system than can reduce fratricide, detectability, and probability of being attacked, as well as minimize system damage, soldier injury, and cognitive and physical fatigue.

(Source: http://www.manprint.army.mil/manprint/additional.html)

> assessments, as well as test and evaluation. These factors interact with each other and with the design and should not be considered separately (p. 38).

The science board also addressed the larger issue of integrating organizational efforts to better support total system development:

> The project manager (PM) must be held responsible for building a system which can perform its specified mission when operated and maintained by well led soldiers. The combat developer must be charged with ensuring that the system is matched to the soldier the Army is likely to access, train, sustain, and retain. These <u>mutually supportive</u> obligations must be bonded in an interactive fashion which, in turn, yields the necessary total system perspective (See Figure below) [emphasis in original] (pp. 38–39).



Finally, the science board bluntly concluded that "integration of personnel acquisition with weapon acquisition is a must" (p. 30) and offered the following broad recommendations:

- Institute a single HMPT authority equal in weight to the materiel authority throughout the system design and decision process

- Focus and resource Soldier research to improve total system performance
- Establish HMPT initiatives with staying power in Army organizations and processes.

By and large, the 1984 Summer Study proved quite favorable in making the case for the change sought by the Office of the DCSPER.

However, even before the Army Science Board completed its study, the Army decided that changes would be made. In June 1984, General Thurman tasked Lt. General Elton and the Office of the DCSPER to review various aspects of HMPT in the Army. The Office of the DCSPER was given staff responsibility for developing policy and clearly defining the HMPT responsibilities of other organizations involved in the acquisition and combat development processes (GAO, 1985, p. 8). That same month, General Richard H. Thompson, a logistics officer, assumed command of the Army Materiel Command. This change afforded General Thurman the opportunity, through General Thompson, to directly inject interest in personnel considerations within the Army's materiel acquisition community (Blackwood & Riviello, 1994, p. 12). General Thompson agreed with Thurman's thrust, although his line of thinking was to institutionalize HMPT integration primarily within the management of the Army Materiel Command (Blanchard & Blackwood, 1990).

Accordingly, in July 1984, General Thompson advanced his idea for "MANPRINT," an acronym he is credited with creating that stood for "manpower and personnel integration" (MANPRINT Policy Office, July 1986, p. 1). The acronym was used both as a noun, to denote the Army Materiel Command's general action plan for imposing HMPT considerations across the materiel acquisition process, and as a verb, in the sense of putting man's imprint or fingerprint on developmental systems (Tragesser, 1985, p. 4). To realize his vision for MANPRINT, General Thompson established a Human Factors Engineering Task Force, comprised of personnel from various Army Materiel Command elements, to analyze HMPT deficiencies, develop a corrective action plan, and oversee implementation of the plan. The task force reviewed previous studies and current policy documents to develop a list of HMPT problems. Field locations were

then visited to verify that the problems still existed. By late August 1984, the task force had developed a matrix of 150 major HMPT action items and a series of tasking letters to be sent to the 31 Army organizations involved in HMPT. These letters delineated those major tasks that fell within each organization's area of responsibility and asked each organization to develop more details of how it planned to perform those tasks. The intent of the task force was to build an Army-wide master plan based on the individual organizational plans. The master plan was to be reviewed and approved by the Army Materiel Command's Materiel Acquisition Review Board in March 1985 (GAO, 1985, p. 9).

Similarly, TRADOC formed a MPT steering committee in September 1984 to develop its own master plan for MPT integration. Its steering committee included officials from several TRADOC organizations who worked to develop a draft master plan by October 1984 (GAO, 1985, p. 10). Seeking to minimize redundancy, General Thompson soon formed a joint Army Materiel Command-TRADOC general officer steering committee to coordinate these two organizations' efforts. Meanwhile, various other smaller commands and agencies, such as the Health Services Command, the Army Safety Center, and the Operational Test and Evaluation Agency, were initiating their own planning activities. Differing perceptions soon developed, and by the fall of 1984 the Office of the DCSPER had come to view the activities of the various commands and agencies with a degree of suspicion. For the Office of the DCSPER, these parallel planning efforts were critical in raising awareness and maintaining the serious attention of the Army's senior leadership on the HMPT problem. However, if history was any guide, many of the initiatives would not have staying power in the Army's organizations and processes given their largely short-term focus (Blackwood & Riviello, 1994, pp. 12–13). The upshot of this dilemma was that the Office of the DCSPER advanced its own comprehensive plan for MANPRINT, which effectively absorbed and subsumed the Army Materiel Command effort. Buoyed by the results of the Army Science Board's 1984 Summer Study, the Office of the DCSPER plan for MANPRINT was briefed to a newly formed General Officer Steering Committee and approved in December 1984 (Blackwood & Riviello, 1994, p. 13).

Unlike the other early initiatives to correct HMPT problems, the Office of the DCSPER plan was based on a strategy for change and represented a careful and deliberative planning process that was focused on a long-term (i.e., 10 year) approach (Blackwood & Riviello, 1994, p. 7). The plan was carefully crafted around a broad set of goals to establish a unifying sense of purpose while still getting to what the Office of the DCSPER viewed as the true crux of the HMPT problem—namely, the limited availability and high cost of personnel resources. The primary challenge for the Office of the DCSPER was the prevailing view within the Army at large that human resource problems were supply problems that should be addressed through supply-side interventions such as enhanced recruiting. Consequently, directly advocating for demand-side interventions on the part of the force planning and materiel acquisition communities was not thought to have a high likelihood for success. To make the Office of the DCSPER concept for MANPRINT marketable, three goals, in a specific order, were put forth: 1) improve human performance, thereby improving total system performance; 2) improve manpower and personnel utilization within the Army; and 3) improve unit effectiveness and readiness by designing and building weapon systems that were easy to use, maintain, and support. Lt. General Elton made the decision to put his staff's primary goal of improving personnel utilization in the middle to deliberately reduce its visibility. Elton believed that the leading and flanking goals, while addressing byproducts of improved personnel utilization, would resonate more directly with field commanders, thereby engendering wide support for their MANPRINT concept (Blackwood & Riviello, 1994, p. 7). As it eventuated, the strategy worked and the first and third goals are now widely appreciated as primary goals of the Defense Department HSI program (DoD, 2008).

There was another calculated logic behind the Office of the DCSPER's choice of goals for their concept of MANPRINT. As mentioned earlier, there was a continual friction between the Office of DCSPER and the ILS community over MANPRINT, which had the unfortunate effect of slowing implementation of, and at times threatening, the very existence of the program. The ILS community was repeatedly shown that MANPRINT could raise supportability issues earlier in the materiel acquisition process, thereby impacting system design so as to head off supportability problems later in the

system life cycle. By implication, MANPRINT and ILS were totally compatible and mutually supportive—at least from the perspective of the Office of the DCSPER. The ILS community, on the other hand, would have none of it. They remained relatively intransigent in their view that MANPRINT was incompatible and redundant to ILS. In the end, this conflict was largely side stepped by emphasizing that the focus of MANPRINT was on performance and not supportability (Blackwood & Riviello, 1994, pp. 4, 20). However, if MANPRINT was to help achieve General Wickham's design objectives for the AOE, it was going to have to address weapon system supportability.

The Office of the DCSPER implementation plan for MANPRINT focused on six key areas that were determined to be essential for institutionalizing change within the Army: policy and procedures, marketing and communications, training and education, resources, research and studies, and evaluations and applications. Specifically, policy and procedures focused on publishing a MANPRINT regulation, changing related existing Army regulations governing the acquisition and combat development processes, and supporting major commands in implementing follow-on guidance. Additionally, MANPRINT positions and skill identifiers were codified in the personnel bureaucracy. Marketing and communications included visits to industry, presentations at professional seminars and conferences, publication of articles, and outreach efforts aimed at the Army's civilian leadership, the other military services, the Office of the Secretary of Defense, and Congressional staffers. Training and education was critical to explaining roles and imparting skills as well as in obtaining active and meaningful involvement from MANPRINT stakeholders. Hay Systems was contracted to support the development and delivery of three nested training courses that were offered starting in January 1986: a 3-week entry level course for MANPRINT practitioners, a 1-week course for supervisors, and a 1-day course for general officers and senior executive service civilians. Resources were required to support program initiatives, but the overall funding strategy was designed to support a minimum program within the Department of the Army (i.e., less than 50 new positions) and tie MANPRINT costs directly to each new acquisition program. Research and studies focused on developing the tools, techniques, and methods for analysis, modeling and simulation, and data management (Blackwood & Riviello,

1994, pp. 13–14). One of these early research initiatives involved the conversion of the Navy's HARDware versus MANpower (HARDMAN) analysis technique for use by the Army to better identify MPT requirements (GAO, 1985, p. 7). Finally, the evaluation and applications initiative focused on gaining feedback from MANPRINT applications and reviews to support an organizational learning cycle (Blackwood & Riviello, 1994, p. 14).

Overall, planning was done across a wide front with the intent of facilitating broad ownership of the program and developing redundancies should certain aspects fail. The eventual success of this approach, particularly with regards to the issue of ownership, can still be witnessed today in response to the question of who started MANPRINT: nearly everybody reports they had a key role in starting MANPRINT. However, the personal interest of Lt. General Elton in "marketing" the program was critical to ensuring that MANPRINT had a highly visible organizational champion in its early years (Blackwood & Riviello, 1994, p. 13). Additionally, the Office of the DCSPER worked to portray the MANPRINT concept as both intuitively simple and coherent with organizational values (Blackwood & Riviello, 1994, p. 9). The simplifying notion played upon the Army's quintessential philosophy about soldiers and machines, namely that the Army equips its men rather than mans its equipment—an assertion that is attributed to General Creighton Abrams circa 1974 (Tragesser, 1985, p. 4). This, in turn, led to the slogan: "MANPRINT: Remember the Soldier." The idea engendered by the slogan was both basic to the Army's culture and struck at the core of the Army leadership principle of "taking care of your soldiers." This had the effect of allowing Lt. General Elton and the Office of the DCSPER to seize the proverbial moral high ground (Blackwood & Riviello, 1994, p. 9).

The Office of the DCSPER decided to primarily focus their efforts during 1985 on building awareness of the MANPRINT concept within the Army (Blackwood & Riviello, 1994, p. 9). Nevertheless, the Office of the DCSPER staff never lost sight of what they thought were the critical aspects of implementing MANPRINT: involvement and communications, training, impacting the acquisition decision process, and influencing source selection (Blackwood & Riviello, 1994, pp. 15–18). Involvement and

communication with those both within and external to the Army was credited with accounting for much of the initial success in institutionalizing MANPRINT. As previously mentioned, a contractor, the Hay Systems, was hired in the summer of 1984 to develop MANPRINT training programs and support program implementation. At the same time, an Army study advisory group, led by the Office of the DCSPER, was formed to assist Hay Systems' work. The advisory group was comprised of over 20 members who represented the MANPRINT domains, the major subordinate commands, and various field agencies. Through the task of developing the MANPRINT training course, the contractor and the study advisory group initially defined and then evolved the MANPRINT roles of the various domain specialists. The study advisory group met monthly for over a year to resolve issues as they emerged. Issues that could not be resolved within the body of the study advisory group were referred to an informal council of colonels, or even to a general officer when appropriate, for solution (Blackwood & Riviello, 1994, p. 15).

In the fall of 1984, the Office of the DCSPER dispatched a fact-finding team comprised of members from the Army Staff, the Army Materiel Command, TRADOC, and the Army Research Institute to visit academia and then industry, starting with those who had representatives on the Army Science Board's summer study. The purpose of these visits was to both facilitate involvement and communication with key external Army stakeholders and to learn how personnel issues were being considered in the engineering design process (Blackwood & Riviello, 1994, p. 15). It was evident from those visits that industry was not playing an active role in addressing personnel issues. However, this was attributed, in large part, to the Army's failure to provide essential information in requirements and contractual documents or to stress consideration of personnel issues in requests for proposals. They concluded that industry had no incentive to engineer designs that would allow the Army to realize personnel savings, especially in system maintenance and support. Thus, these visits highlighted the need for MANPRINT to focus on reliability, availability, and maintainability issues as they affected maintenance and support personnel considerations (W. O. Blackwood, personal communication March 3, 2010). Concurrent to the work of the fact-finding team, Hay

Systems set up a series of three meetings between General Thurman and senior executives from several large defense firms, with each meeting involving managers from different firms. These meetings provided a venue for frank and open discussions at the senior executive level and served to validate the observations of the fact-finding team. Such meetings were held at Fort McNair and continued for several years as a means for soliciting inputs and communicating progress and expectations (Blackwood & Riviello, 1994, pp. 15–16).

Beyond the outreach efforts, the spring of 1985 saw the publication of the draft version of Army Regulation 602-2, *Manpower and Personnel Integration (MANPRINT) in the Material Acquisition Process*, clarifying MANPRINT roles and responsibilities. Meanwhile, the Army Materiel Command and TRADOC had revised program managers' and TRADOC system managers' statements of responsibility to more clearly articulate their accountability for MANPRINT in the materiel acquisition process (W. O. Blackwood, personal communication March 3, 2010). Given the complexity of the materiel acquisition process and the sheer quantity of documents produced, the Office of the DCSPER chose to focus initial efforts on three strategic points in the weapon system acquisition process: the request for proposal (RFP), source selection, and the test and evaluation process (Blackwood & Riviello, 1994, p. 17).

The RFP is a formal solicitation by the government for industry to provide proposals for a specific commodity or service. It was clear from the observations of the fact-finding team that industry would address MANPRINT considerations if they were a requirement in the RFP. Thus, a major initiative by the Office of the DCSPER was to ensure that RFPs incorporated MANPRINT tasks in the statement of work, included deliverables for those tasks, and identified MANPRINT considerations as a factor in source selection (Blackwood & Riviello, 1994, p. 17).

Source selection is the process by which the government evaluates responses by industry to a RFP and selects a winning proposal. As it turned out, establishing a policy that caused MANPRINT to be included as a major element in the source selection criteria was one of the most difficult challenges faced by the Office of the DCSPER (Blackwood & Riviello, 1994, p. 17). Although Army Regulation 602-2 (Department of the Army,

1987) provided the formal direction that the "MANPRINT assessment will be a separate major area in source selection and evaluations" (p. 3), the materiel acquisition community was very resistant given their strong belief that this would increase program costs. The basic supposition that MANPRINT would actually reduce total ownership costs was never an issue—opponents of MANPRINT were primarily focused on the issue of procurement costs. The real controversy was the fact that the inclusion of MANPRINT in source selection would directly impact industry, obligate resources, and limit the ability of the materiel acquisition community to waive MANPRINT requirements. The source selection issue was sufficiently cantankerous that it had to eventually be resolved at the Under Secretary of the Army level through the publication of Army Acquisition Executive Policy Memorandum 89-2, which directed that MANPRINT be both included in all RFPs and considered a factor in source selection evaluations. Unfortunately, the Army Materiel Command, which was adamantly opposed to including MANPRINT in source selection, was responsible for enforcing the source selection policy, an arrangement that almost guaranteed that the controversy would be perpetuated. In the long run, however, this situation actually contributed to the health of the MANPRINT program by continuing to force the Army's senior leadership to study the issue as conflicts resurfaced (Blackwood & Riviello, 1994, pp. 17–18, 21).

Since MANPRINT was to be applied across the materiel acquisition process, it was imperative that a mechanism be provided for influencing acquisition decision making. This was achieved by making the Office of the DCSPER an effective participant on the Army's top-level materiel acquisition decision board, the Army Systems Acquisition and Review Council (ASARC). The DCSPER representative to the ASARC could ensure that critical human performance issues were at least identified and acknowledged before a final decision on design, procurement, test, or fielding was made by the council. Accordingly, it became the ASARC's *de facto* role to evaluate whether MANPRINT was effectively applied to a system under review. This situation had the distinct advantage that the materiel acquisition community perceived that they were answering for MANPRINT issues to the Army's highest acquisition authority and not the Office of the DCSPER. This helped to diffuse ownership for MANPRINT throughout

the materiel acquisition community while providing MANPRINT a significant go/no-go vote on every major Army acquisition program (Blackwood & Riviello, 1994, p. 21).

During 1985, a number of pilot projects were selected to expedite learning and gain experience with implementing MANPRINT in each of the major materiel acquisition phases. But, it was the Light Helicopter Experimental (LHX) program, or what became known as RAH-66 Comanche, that was the key pilot program for demonstrating the viability and potential effects of MANPRINT. The LHX program was specifically chosen because it was the Army's largest and most visible program at the time and the program manager was supportive of MANPRINT. Thus, the LHX program became the first major Army program to both implement MANPRINT considerations into the front end analysis phase of the materiel acquisition process and to include MANPRINT in the source selection document (MANPRINT Policy Office, July 1986). The LHX program would prove crucial in establishing the credibility of the MANPRINT effort, and hence, Lt. General Elton took significant personal interest, chairing the LHX MANPRINT review (W. O. Blackwood, personal communication March 3, 2010). By 1986, LHX became a true experimental program, testing where it was possible to introduce advanced technology into the Army's inventory without creating problems of unsatisfactory total system performance or increasing personnel demands. Even opponents of the LHX program were impressed by the advances achieved relative to the standard of normal acquisition practices. It was later estimated in 1995 that the potential cost avoidance in the LHX program in terms of manpower, personnel, training, and safety was $3.3 billion, equating to an 8,000 percent return on investment for the portion of the program's research and development budget that was attributable to MANPRINT (Skelton, 1997). Other successful early applications of MANPRINT included the pedestal-mounted Stinger missile system, the line of sight-forward heavy (LOS-H) air defense artillery system, and the howitzer improvement program (HIP) (Booher, 1988, p. 2).

By the fall of 1985, Lt. General Elton recognized that the growing MANPRINT effort would soon require more attention than he could realistically provide given his other duties as DCSPER. Additionally, he was cognizant of the finite time that he had

left in his current assignment and appreciated the importance of continuity of leadership to the health of the program. Consequently, a civilian senior executive service (SES) position was created within the Office of the DCSPER to serve as the director for the Army's MANPRINT program. While the overall MANPRINT staff would be kept relatively small, the fact that it was led by a SES helped ensure it had appropriate visibility and authority (Blackwood & Riviello, 1994, p. 10). Moreover, the SES would wield a big stick as the Office of the DCSPER's representative on the ASARC. In July, 1986, Dr. Harold R. Booher was hired on as the first civilian director of the Army's MANPRINT program (MANPRINT Policy Office, August 1986). With Booher at the helm as a Special Assistant to the DCSPER, the MANPRINT office, which had started in the DCSPER's Research and Studies Office, became first a Special Assistant Office in 1986 and then an official directorate in the Office of the DCSPER in 1987.

4. Human Systems Integration (1987–)

The MANPRINT program was designed to survive on its own after 1986, although there were serious doubts at the time as to whether it would. With Booher in the MANPRINT Office to provide continuity and leadership, and with General Thurman able to continue his advocacy through 1989, first as Vice Chief of Staff of the Army and then as commander of TRADOC, MANPRINT was afforded sufficient attention and support from 1986 through 1989 to maintain momentum institutionalizing the program within the Army. In October 1987, Lieutenant Colonel Blackwood, who had been selected in 1983 by Lt. General Elton to help plan, organize, and implement the MANPRINT program, was tapped to join the Strategic Planning Office of the Under Secretary of Defense for Acquisition. Now Blackwood would help instigate a similar effort within the Office of the Secretary of Defense (Blackwood & Riviello, 1994, pp. 7, 19, 21-22; W. O. Blackwood, personal communication March 3, 2010).

Before continuing with this historical narrative, we must backup to introduce an important supporting actor in the story of how MANPRINT came into being—one Mr. Delbert L. Spurlock, Jr. Mr. Spurlock was a lawyer by trade and had come from private practice to serve as the General Counsel at the Department of the Army starting in 1980.

Spurlock was appointed Assistant Secretary of the Army for Manpower and Reserve Affairs in 1984, making him the highest Army civilian representative for the personnel community, and not surprisingly, a strong advocate for Lt. General Elton's MANPRINT concept. Spurlock's interest and support mainly stemmed from an Army manpower perspective, but he also had a vision that extended the MANPRINT philosophy throughout the military and beyond, into the nation's workforce at large. But at the same time, Spurlock was skeptical about whether the Army would stick to its MANPRINT master plan. Spurlock testified before the House Armed Services Committee to this effect in the spring of 1985. As related by Booher (1988):

> In his testimony he explained that weapons developers have not been interested in human factors implications because they were "perceived as constraints" on producing systems "on time" within "cost." Moreover, to expect accountability and wise decision-making with MANPRINT "through one of the world's largest bureaucracies" without the support of Congress would probably be wishful thinking. With history as a guide, MANPRINT too would fail. Spurlock's testimony made clear that MANPRINT could work, however, if Congress insisted that "total system manpower quality and quantity, and training costs were prerequisite findings for any weapon system beyond concept exploration" (p. 2).

According to Spurlock (as quoted in Patrick, 1988):

> I recommended that Congress get more involved in the process…and require some sort of analysis from the Services that quantified the MPT (manpower, personnel and training) burden. That is essentially what led to the Manpower Estimate Report (p. 5).

As it was established, the Manpower Estimate Report (MER) documented the human costs associated with operating and supporting a weapon system throughout its life cycle. The MER process was statutorily imposed on the Defense Department by the Fiscal Year 1987 Defense Authorization Act, which required that a MER be submitted to Congress by the Secretary of Defense prior to his approval of full-scale development and/or production and fielding of a major weapon system or program that had Congressional interest. These requirements were codified in Title 10, United States Code, Section 2434 (U.S. Congress, 1986), and the Defense Department formally implemented them in DoD

Directive 5000.53, titled *Manpower, Personnel, Training and Safety in the Defense System Acquisition Process*, in December 1988 (Bergquist, 1991, p. 4).

From Spurlock's perspective, the MER requirement that was imposed across the military services was, in effect, a Congressional mandate to address MANPRINT considerations (Spurlock as quoted in Patrick, 1988):

> MANPRINT is the heart of the MER process. You cannot quantify the kinds of operating and supports costs we are talking about without first performing a MANPRINT kind of analysis (p. 6).

Although Spurlock saw similarities between MANPRINT and the MER process, he and others continued to work to see the MANPRINT concept more fully embraced by the Defense Department. According to Spurlock, both he and "a number of our young officers" were in "day-to-day contact with their counterparts in [the Office of the Secretary of Defense]" helping to "sell the concept" of MANPRINT (Spurlock as quoted in Patrick, 1988, p. 6). Spurlock also described a more direct conversation with Mr. Richard Godwin, the Under Secretary of Defense for Acquisition, in which he "spent an hour or so talking about MANPRINT and why [Mr. Godwin] should make this a requirement for all of the Services" (Spurlock as quoted in Patrick, 1988, p. 6).

While Spurlock was providing high level MANPRINT advocacy on the part of the Army's civilian leadership, Blackwood had begun pushing MANPRINT concepts from within the Office of the Secretary of Defense soon after his arrival in 1987. He was helped in this matter by a senior civilian defense executive, that being Mr. Thomas Christie—Boyd Acolyte[14] and member of the inner circle of the military reform movement. While Christie was never to become an outright MANPRINT advocate *per se*, he did see MANPRINT concepts as being complementary to the core views held by the reformers. Accordingly, Christie was supportive of Blackwood's efforts to advance MANPRINT concepts within the Office of the Secretary of Defense and provided assistance in terms of access to both his formal and informal networks within the Defense

[14] Christie, along with Pierre Sprey, Ray Leopold, Franklin Spinney, and Jim Burton were described by writer Robert Coram as (Colonel John) Boyd's Acolytes, a group who, in various ways and forms, promoted and disseminated Boyd's ideas throughout the modern military and defense establishment.

Department. Although Blackwood was now technically working within the defense acquisition community, he firmly believed, given his experience in the Army, that the personnel community was the preferred organizational sponsor for a Defense Department-level MANPRINT initiative. Over the ensuring year, he worked with his counterpart in the Office of the Assistant Secretary of Defense for Force Management and Personnel (ASD/FM&P), Air Force Lt. Colonel Michael Pearce, to convince the personnel community to take the lead in advocating for the MANPRINT concept. Blackwood's early overtures received a tepid reception by ASD/FM&P, at least until it became known that the sponsorship of the acquisition executive was also being solicited.

Through the persistent efforts of MANPRINT advocates like Mr. Spurlock and Lt. Colonel Blackwood, ASD/FM&P was duly persuaded to signal the Defense Department's organizational commitment to MANPRINT goals. Consequently, in addition to the MER process, DoD Directive 5000.53 established manpower, personnel, training, and safety (MPTS) criteria that were required to be addressed by all the military services in cooperation with industry:

> The Department of Defense shall maximize the operational effectiveness of all systems, whether being procured initially or being refurbished, by ensuring those systems can be effectively operated, maintained, and supported by well qualified and trained people. To do so, human capabilities and limitations must be fully considered early in the system design process. Such MPTS concepts, requirements and goals shall be developed in a consistent manner, communicated to industry, evaluated in contract proposals, and weighed positively and substantially as criteria for source selection (DoD Directive 5000.53 as quoted in Boff, 1990).

It also required that MPTS considerations be assessed, documented, and reported to ASD/FM&P at each phase of the acquisition process. Noteworthy in DoD Directive 5000.53 was the inclusion of MPTS considerations as a major element in the source selection criteria, something that had been adamantly opposed by the Army's materiel acquisition community during the initial implementation of MANPRINT. Equally noteworthy was the omission of human factors engineering from among the criteria—a measure that proved necessary to circumvent an impasse that had arisen between the service representatives comprising the conference responsible for drafting the DoD

directive.[15] Nevertheless, DoD Directive 5000.53 was considered a landmark accomplishment in the evolution of the Defense Department's HSI program.

During this same period, a nascent HSI office was established within the Office of the Assistant Secretary of Defense for Force Management and Personnel (OASD/FM&P) with Lt. Colonel Pearce as the chief. At the time, there was a certain intuitive logic to the placement of HSI within OASD/FM&P given that training and readiness were in the assistant secretary's mission statement. Additionally, manpower and training were major cost drivers for the Defense Department and personnel budgets were a primary concern of OASD/FM&P. However, as time passed, this arrangement quickly proved a less than ideal marriage (P. Chatelier, personal communication May 25, 2010).

Given its organizational location, the OASD/FM&P HSI office primarily concerned itself with training and readiness issues.[16] Accordingly, it reviewed weapon system acquisition programs mainly in terms of manpower requirements, readiness documents, and training documents. Representatives from the office also sat on a variety of committees within the individual Services but generally lacked sufficient organizational clout to effect decisions that were needed. Part of the problem was a personnel issue: the office was led by a lieutenant colonel, which was an insufficient rank given the power gradients of the Pentagon. In addition, the OASD/FM&P bureaucracy failed to provide sufficient personnel and budgetary resources to grow the HSI office. Nor did it have access to research, development, test, and evaluation

[15] A significant source of friction was caused by varying perceptions regarding the nature of "human factors." At the time, human factors was primarily considered the province of those working in the life sciences—a notion that did not resonate well with the Services' engineering communities. Accordingly, the engineering communities, which had responsibility for human-machine interface design, sought to carve a distinction between the engineering-related elements of human factors (i.e., human factors engineering) and those related to manpower, personnel, training, and education. Thus, the omission of human factors from DoD Directive 5000.53 was indeed deliberate.

[16] One well-documented initiative sponsored by the OASD/FM&P HSI Office was the DoD Liveware survey, which was an attempt, in conjunction with the North Atlantic Treaty Organization (NATO) Research Study Group.21 (RSG.21), to survey the HSI community to obtain a comprehensive database of HSI technologies. In addition to serving as chief of the OASD/FM&P HSI office, Pearce chaired RSG.21, designated *Liveware Integration in Weapon System Acquisition*. RSG.21 was chartered by the NATO Defense Research Group 8, *Defense Applications of Human and Bio-medical Sciences*, to study how the human-machine interface was addressed and how human-related issues were resolved during acquisitions by member nations. It was RSG.21 that coined the term "liveware" in an attempt to standardize references to HSI across national boundaries and languages (Gentner, Kancler, & Crissey, n.d.).

(RDT&E) or procurement monies, these being the official currencies of the acquisition community. These issues—or more poignantly, the failure to correct them—were reflective of a general resistance on the part of the OASD/FM&P to getting involved in equipment-related issues and system acquisition. Simply stated, senior personnel within OASD/FM&P could not make the conceptual leap from human factors engineering and logistics considerations within individual weapons systems programs to aggregate concerns about Defense Department manpower and personnel. Thus, they felt obliged to let the acquisition community drive research and development initiatives. Even in the area of training, which was a focus area for Pearce, the emphasis was primarily on advanced distributed learning, with issues of training hardware and simulation being deferred to the Office of the Director of Defense Research and Engineering. Not surprisingly, given the organizational atmospherics of the time, the OASD/FM&P HSI office quickly faded with the departure of Pearce (P. Chatelier, personal communication May 25, 2010).[17]

Fortunately, Defense Department level HSI policy guidance did not meet the same fate as the OASD/FM&P HSI office. Instead, it appears HSI policy guidance followed an independent evolutionary trajectory, the story of which necessitates that we again backtrack for a moment. In July 1989, Secretary of Defense Richard B. Cheney delivered his Defense Management Report, which outlined how the Defense Department would implement the Packard Commission's recommendations to streamline the materiel acquisition process, increase tests and prototyping, change the organizational culture, and improve planning, among other things (Cheney, 1989). The Defense Management Report was largely critical of materiel acquisition management, describing it as both undisciplined and overburdened by regulation. As a result, Deputy Secretary of Defense Donald J. Atwood issued the 1991 edition of DoD Directive 5000.1, titled *Defense Acquisition*, and DoD Instruction 5000.2, titled *Defense Acquisition Management Policies and Procedures*. These documents were expanded from the prior versions to

[17] No one that was interviewed could recall the specific year at which the OASD/FM&P HSI office was officially dissolved. The general consensus was that the office became defunct sometime during 1993-1994.

absorb over 60 other directives, instruction, and memoranda to include DoD Directive 5000.53 (U.S. Army Center of Military History, 2008).

This revision of the DoD 5000 series of documents replaced MPTS requirements with those addressing Human Systems Integration (HSI). This was, however, more than simply a name change.[18] Greater definition was provided regarding the components that comprised HSI and increased discipline was required in the documentation process that connected requirements and opportunities with resource decisions (Hewitt, 1991, p. 6). These directives included major sections on HSI; systems engineering/human factors engineering; integrated logistics support/manpower, personnel, and training (MPT); and systems safety, health hazards, and environmental impact. The HSI section in particular stated the policy objective as follows: "Human considerations [Figure III-11] shall be effectively integrated into the design effort for defense systems to improve total system performance and reduce costs of ownership by focusing attention on the capabilities and limitations of the soldier, sailor, airman, or marine" (DoD, 1991, p. 7-B-1). Additionally, it directed that "objectives for the human element of the system shall be initially established at Milestone I…and be traceable to readiness, force structure, affordability, and wartime operational objectives" (DoD, 1991, p. 7-B-1). Furthermore, the revision effectively moved HSI from an ASD/FM&P owned directive to one controlled by the Under Secretary of Defense for Acquisition despite the fact that the Defense Department's HSI office resided in the Office of the ASD/FM&P.

[18] It is uncertain who first coined the term "human systems integration." William Blackwood, who helped initiate the Army MANPRINT program and the Defense Department HSI program, believes the term was developed by someone working for the Air Force in Dayton, Ohio. Blackwood recounts that he was interested in changing the name of the Army's nascent program in the 1980s from MANPRINT to a more gender-neutral term. His thinking at the time was that MANPRINT was seemingly incongruent with the Army's ongoing efforts to increase the recruitment of women as part of the all-volunteer force. However, given the strong headwind Blackwood was facing in just getting MANPRINT off the ground, it was decided that a name change was too low a priority given all the other competing demands for time.

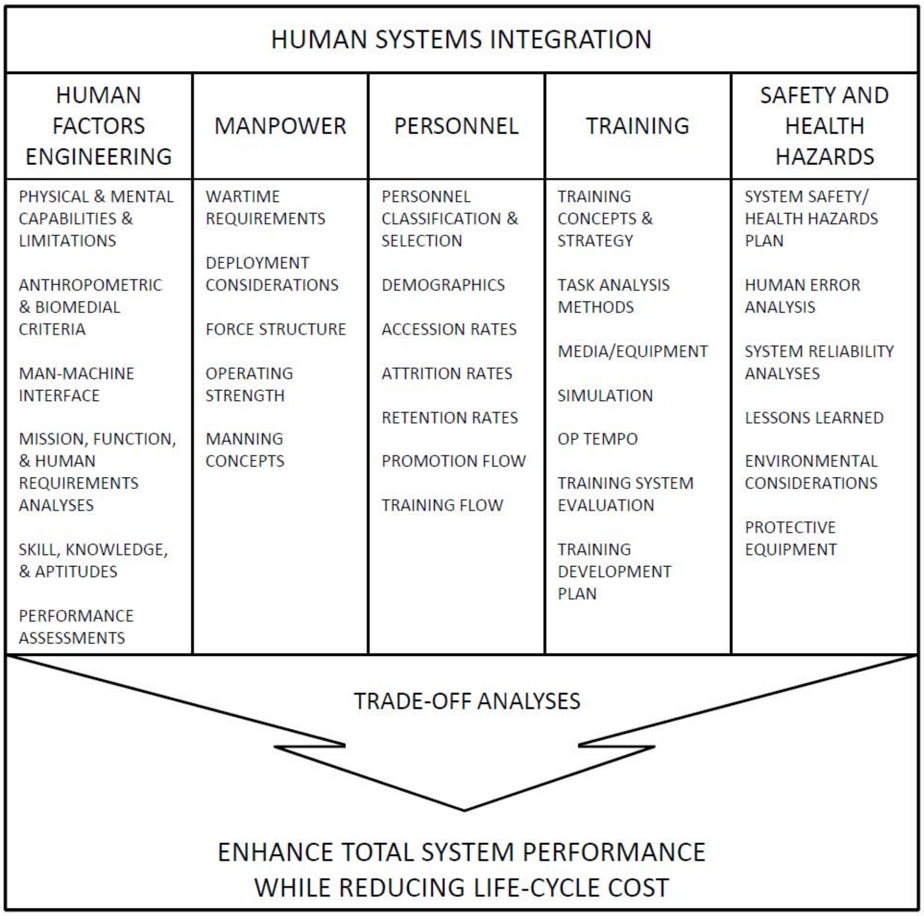


Figure III-11. Depiction of the human considerations addressed in HSI as described in the 1991 edition of DoD Instruction 5000.2 (From DoD, 1991).

Starting with the 1991 edition of DoD Directive 5000.1 and DoD Instruction 5000.2, HSI has remained a component of the DoD 5000 series of acquisition management policy documents. Subsequent editions of DoD Instruction/Regulation 5000.2, released in 1996, 2000, 2002, 2003, and most recently 2008, continued to feature one or more sections addressing HSI. The HSI section in the 1996 and 2000 editions of DoD Regulation 5000.2-R were much diminished from that in the 1991 edition, describing the policy objectives for HSI as ensuring that “human performance; the burden the design imposes on manpower, personnel, and training (MPT); and safety and health aspects are considered throughout the system design and development processes” (DoD, 1996 & 2000, part 4, p. 8). In the 2002 edition of DoD Regulation 5000.2-R, the stated

policy objectives for HSI were revised to focus on the need to "optimize[19] total system performance and minimize [total ownership costs]" by integrating "manpower, personnel, training, safety and occupational health, habitability, human factors, and personnel survivability considerations" (DoD, 2002, pp. 40–42, 98–99). The 2003 edition of DoD Instruction 5000.2 established the contemporary policy verbiage for HSI, namely to "optimize total system performance, minimize total ownership costs, and ensure that the system is built to accommodate the characteristics of the user population that will operate, maintain, and support the system" (DoD, 2003, p. 44). It further defined HSI as being comprised of the following components: human factors engineering; personnel; habitability; manpower; training; environment, safety, and occupational health (ESOH);

[19] The use of the word "optimize" starting in the 2002 edition of DoD Instruction 5000.2-R has led to some ambiguity as to the overall objective of the Defense Department HSI program. Given the historical legacy of systems analysis and operations research in the Defense Department, it is reasonable to consider the term "optimize" in its mathematical programming sense. The issue then becomes one of identifying the objective or criterion function that is to be optimized.

Since William Blackwood was principally responsible for authoring the initial policy that laid out the formal definition of MANPRINT (i.e., AR 602-2) and later helped instigate the Defense Department level implementation of HSI, the optimization question was specifically put to him during an interview on March 3, 2010. His response was that the use of the word "optimize" should be considered in terms of minimizing the following criterion: the change in observed performance per unit aptitude ($\Delta P \div \Delta A$). As an example, Blackwood used the comparison of the M60 and M1 tank crew performance described in Binkin (1986). The M1 Abrams, the Army's newest tank in the early 1980s, was equipped with a full-solution fire control system featuring a laser range finder, ballistic computer, thermal-imaging night gunsight, full stabilization, a muzzle reference system to measure gun-tube distortion, and a wind sensor. It was clearly more complex than its predecessor, the M60 tank. The table below, taken from Binkin (p. 55), arrays range firing results by crews manning M1 tanks with those manning the older M60. M1 crews consistently scored more tank "kills" than M60 crews with similar aptitudes. But more interestingly, the number of kills by M1 crews was less influenced by their aptitudes, suggesting that the more complex M1 system was easier to employ than its predecessor and could be successfully operated by crew members with lower aptitude scores. In essence, the Army could relax aptitude requirements for tank crews, thereby increasing the availability of potential crewmembers. The end result was improved utilization of Army personnel, which was the primary goal of the MANPRINT program.

AFQT category of gunner/tank commander	*Tank equivalent kills*		
	M60	*M1*	*Percent improvement*
I (above average)	10.23	12.75	25
II (above average)	9.51	12.47	31
IIIA (average)	8.52	12.05	41
IIIB (average)	7.47	11.57	55
IV (below average)	5.84	10.72	84

and survivability. In the 2008 edition of the 5000 series, the components that comprised HSI were slightly modified by the deletion of "environment" (DoD, 2008, pp. 60–61).

E. CONCLUSION

So what does the historical narrative tell us about HSI? HSI appears to have first emerged as a result of the spread of the systems approach, and particularly systems analysis, from the RAND Corporation to the Defense Department. In the U.S. Army's Human Engineering Laboratory in the 1960s, John Weisz and colleagues worked to join human factors engineering and operations research to more broadly represent human considerations in weapon system analyses. This was the origin of the conceptual framework for what would become HSI. Unfortunately, the initial launch of HSI, occurring during the Army's "lost decade" for materiel development, resulted in a resounding thud.

HSI would not take off until the Army underwent a doctrinal and organizational renaissance in the late 1970s and early 1980s, driven in large part by fears of an apocalyptic war with the Soviet Union on the plains of Europe. This led to another rise of science-based military power as the Army sought to leverage high technology to achieve a credible parity with the numerically superior Soviet forces. But, in so doing, the Army began to have trouble coping with complexity, particularly with regards to their most politically constrained resource—personnel. A crisis ensued during the design of the Army of Excellence and the need to find personnel to create two new light infantry divisions. The Army's solution was to better utilize their human resources, especially those tied up in the maintenance and support of weapon systems. Thus, the resurrection of HSI occurred in 1983, this time spearheaded by a coalition of advocates who, interestingly enough, were not of the human factors community. These people were senior military and civilian leaders or experts in organizational change; many had unique backgrounds and experiences that made them particularly well suited for the tasks at hand. They developed a systems discourse, centrally including the materiel acquisitions and personnel communities, which was eventually to be institutionalized in the Army's

bureaucracy as MANPRINT. Many of these same people then carried the discourse into the Defense Department bureaucracy, where it became formally codified as HSI.

The moral of this story is that real-world political and military situations drove the adoption of HSI and the elaboration of its constituent components and ultimate goals. To put it rather bluntly, it was a means to achieving an end that was seen as essential to the viability of the Army as a credible defense against the Soviet Union. The lesson of history, then, is that taking a closed or scientific view of HSI as it now exists within the Defense Department is fundamentally flawed. Such a view ignores the open-ended reality of politics and the endless creativity of human beings, both in fighting and resolving organizational conflicts. It remains to be demonstrated whether the application of the systems discourse that surrounded the Army's MANPRINT program would yield a similar program in an entirely different organizational context. My instinct is that it would not.

F. BIBLIOGRAPHY

Army Research Laboratory (2003). *A history of the Army Research Laboratory*. Adelphi, MD: Author.

Bergquist, T. M. (1991). *Air Force manpower, personnel, and training (MPT) in systems acquisition research program* (Report No. AL-TP-1991-0021). Brooks Air Force Base, TX: Human Resources Directorate.

Betaque, N. E., Jr., Kennelly, D. M., Nauta, F., & White, T. A. (1978). *Manpower planning for new weapon systems*. Washington, DC: Logistics Management Institute. (DTIC Accession No. ADA066928)

Binkin, M. (1986). *Military technology and defense manpower*. Washington, DC: The Brookings Institution.

Blackwood, W. O., & Riviello, R. N. (1994). *Organizational change lessons learned from MANPRINT: a case study* (Contract No. MDA-903-91-C-0004). Washington, DC: Army MANPRINT Office.

Blanchard, R. E., & Blackwood, W. O. (1990). Change management process. In H. R. Booher (Ed.), *MANPRINT: an approach to systems integration* (chap. 3, pp. 55−93). New York, NY: Van Nostrand Reinhold.

Boff, K .R. (1990). Meeting the challenge: factors in the design and acquisition of human-engineered systems. In H. R. Booher (Ed.), *MANPRINT: an approach to systems integration* (chap. 19, pp. 551−572). New York, NY: Van Nostrand Reinhold.

Booher, H. R. (1988, Dec). Progress of MANPRINT – the Army's human factors program. *Human Factors Society Bulletin*, 31(12), 1−3.

Booher, H. R. (1990). Introduction: the MANPRINT philosophy. In H. R. Booher (Ed.), *MANPRINT: an approach to systems integration* (chap. 1, pp. 1−16). New York, NY: Van Nostrand Reinhold.

Cheney, Dick (1989). *Defense management report to the president*. Washington, DC: Government Printing Office.

Compston, C. & Seidman, R. F. (Eds). (2003). *Our documents: one hundred milestone documents from the National Archives*. New York, NY: Oxford University Press.

Coram, R. (2002). *Boyd: the fighter pilot who changed the art of war*. New York, NY: Little, Brown, and Company.

Deal, S. (2007, January). *Definition of human systems integration*. HSI working group meeting conducted at International Council of Systems Engineering International Workshop 2007, Albuquerque, New Mexico.

Department of the Army (1966, December). *Report by the Department of the Army Board of Inquiry on the Army Logistics System: Volume III. Acquisition management*. Washington, DC: Author.

Department of the Army (1967, March). *Report by the Department of the Army Board of Inquiry on the Army Logistics System: Vol I. Introduction, summary, and guide to implementation*. Washington, DC: Author.

Department of the Army (1968). *Man-material systems: human factors engineering program* (AR 602-1). Washington, DC: Author.

Department of the Army (1976). *Personnel-material systems: human factors engineering program* (AR 602-1). Washington, DC: Author.

Department of the Army (1987). *Manpower and personnel integration (MANPRINT) in materiel acquisition process* (AR 602-2). Washington, DC: Author.

Department of Defense (1991). *Defense acquisition management policies and procedures* (DoDI 5000.2). Washington, DC: Author.

Department of Defense (1996). *Mandatory procedures for major defense acquisition programs (MDAPs) and major automated information system (MAIS) acquisition programs* (DoDR-5000.2-R). Washington, DC: Author.

Department of Defense (2000). *Mandatory procedures for major defense acquisition programs (MDAPs) and major automated information system (MAIS) acquisition programs* (DoDR 5000.2-R). Washington, DC: Author.

Department of Defense (2002). *Mandatory procedures for major defense acquisition programs (MDAPs) and major automated information system (MAIS) acquisition programs* (DoDR 5000.2-R). Washington, DC: Author.

Department of Defense (2003). *Operation of the defense acquisition system* (DODI 5000.2). Washington, DC: Author.

Department of Defense (2008). *Operation of the defense acquisition system* (DODI 5000.02). Washington, DC: Author.

Dupay, A. P. (1988). *The Army of Excellence: at what price to combat service support?* (Study project). Carlisle Barracks, PA: U.S. Army War College. (DTIC Accession No. ADA195741).

Edwards, P. N. (1988). The closed world: systems discourse, military strategy and post WWII American historical consciousness. *AI and Society*, 2, 245–255.

Edwards, P. N. (1996). *The closed world: computers and the politics of discourse in Cold War America*. Cambridge, MA: MIT Press.

Fischer, H. E., & Fischer, E. (1999). *Editorial cartoon awards 1922-1997: from Rollin Kirby and Edmund Duffy to Hebert Block and Paul Conrad*. München, Germany: Saur.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327–359.

General Accounting Office (1981). *Effectiveness of U.S. forces can be increased through improved weapon system design* (Report No. PSAD-81-17). Washington, DC: Author.

General Accounting Office (1985). *The Army can better integrate manpower, personnel, and training into the weapon systems acquisition process* (Report No. NSIAD-85-154). Washington, DC: Author.

Gentner, F. C., Kancler, D. E., & Crissey, M. J. (n.d.). Results of the DoD liveware survey. *IEEE*. Retrieved June 10, 2010, from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=290865&userType=&tag=1

Guilmartin, J. F. Jr., & Jacobowitz, D. W. (1985, March 21). *Group interaction and the design of military technology: a historical analysis of an interactive process.* Paper presented at Annual Meeting of the Southwestern Social Science Association, Houston, TX.

Hammond, G. T. (2001). *The mind of war: John Boyd and American security*. Washington, DC: Smithsonian Books.

Hawkins, G. R., & Carafano, J. J. (1997). *Prelude to Army XXI: U.S. Army division design initiatives and experiments 1917–1995*. Washington, DC: U.S. Army Center of Military History. Retrieved March 31, 2010, from http://www.fas.org/man/dod-101/army/unit/docs/xxi/toc.htm

Herzog, S. (1994). *Defense reform and technology: tactical aircraft*. Westport, CT: Praeger.

Hewitt, G. (1991, Jan-Feb). Human-systems integration: now more than ever. *MANPRINT Bulletin*, V(3), 6–8.

Holley, Jr., I .B. (2004). *Technology and military doctrine: essays on a challenging relationship*. Maxwell Air Force Base, AL: Air University Press.

Hughes, T. P. (1998). *Rescuing Prometheus*. New York, NY: Vintage Books.

Human Engineering Laboratory (1969 Jan). *Manpower resources integration guide for Army materiel development* (HEL Guide 1-69). Aberdeen Proving Ground, MD: Author. (DTIC Accession No. AD698451)

Illingworth, L. G. [Drawing of Stalin planting Soviet flags in Europe] (1947, June 16). *Daily Mail*. Retrieved January 26, 2010, from http://www.cartoons.ac.uk/record/ILW1262

Illingworth, L. G. (1950). *Communist threats all over the world* [Drawing]. National Library of Wales. Retrieved January 26, 2010, from the online Illingworth Cartoon Exhibition at http://www.llgc.org.uk/illingworth/index_s.htm

Kerwin, W. T., Blanchard, G. S., Atzinger, E. M., & Topper, P. E. (1980). *Man/machine interface – a growing crisis* (Army Top Problem Areas Discussion Paper Number 2). Aberdeen Proving Ground, MD: U.S. Army Material Systems Analysis Activity. (DTIC Accession No. ADB071081)

Kross, W. (1985). *Military reform: the high-tech debate in tactical air forces*. Washington, DC: National Defense University Press.

MANPRINT Policy Office (1986, Jul). LHX MANPRINT implementation. *MANPRINT Bulletin*, 1, 1−2.

MANPRINT Policy Office (1986, Aug). SES appointed to head MANPRINT initiative. *MANPRINT Bulletin*, 2, 1−2.

McNamara, R. S. (1995). *In retrospect: the tragedy and lessons of Vietnam.* New York, NY: Vintage Books.

Nucci, E. J. (1967). *Study of manpower considerations in development. Volume 1. Findings and recommendation of the study group*. Washington, DC: Office of the Director of Defense Research and Engineering. (DTIC Accession No. ADB826438)

Ostovich, R., Folwer, D. J., & Hatlestad, J. S. (1982). *Soldier machine interface requirements* (*"complexity"*) *study. Volume 1. Main report* (Report No. ACN 36879). Fort Leavenworth, KA: U.S. Army Combined Arms Combat Development Activity. (DTIC Accession No. ADB070710).

Popper, K. R. (1957). *The poverty of historicism*. London, U.K.: Routledge and Kegan Paul.

Promisel, D. M., Hartel, C. R., Kaplan, J. D., Marcis, A., & Whittenburg, J. A. (1985). *Reverse engineering: human factors, manpower, personnel, and training in the weapon system acquisition process* (Technical Report No. 659). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC Accession No. ADA157693)

Romjue, J. L. (1993). *The army of excellence, the development of the 1980s army*. Fort Monroe, VA: Office of the Command Historian, U.S. Army Training and Doctrine Command. (DTIC Accession No. ADA269724)

S 35s captured by Germany in 1940 [Photograph]. Retrieved June 6, 2010, from: http://en.wikipedia.org/wiki/File:Bundesarchiv_Bild_121-0412,_Frankreich,_Panzer_ Somua_S35,_Gesch%C3%BCtz.jpg

Sargent, B. (1981, September 10). Defense system [Drawing]. *The Austin American-Statesman*, 111(47), A14.

Sheehan, N. (2009). *A fiery peace in a Cold War: Bernard Schriever and the ultimate weapon*. New York, NY: Random House.

Patrick, R. (1988, Nov-Dec). A view from the top, an interview with the honorable Delbert L. Spurlock Assistant Secretary of the Army for Manpower and Reserve Affairs. *MANPRINT Bulletin*, III(3), 5−7.

Skelton, I. (1997, October 1). *MANPRINT for the U.S. Army*. Statement by Congressman Ike Skelton. Washington, DC.

Spinney, F. C. (1985). *Defense facts of life, the plans/reality mismatch*. Boulder, CO: Westview Press.

Stanley, W. D. III (1985, Mar-Apr). MANPRINT: the leverage for excellence. *Army Research, Development, and Acquisition Magazine*, 5−7.

Stewart, R. W. (Ed.) (2005). The United States Army in a global era, 1917–2003. *American military history* (Vol. 2). Washington, DC: U.S. Army Center of Military History.

Sun Tzu (1963). *The art of war* (S. B. Griffith Trans.). New York, NY: Oxford University Press.

Time Magazine (1957, April 29). *Engineers Woolridge and Ramo* [Cover illustration]. Retrieved January 26, 2010, from http://www.time.com/time/covers/0,16641, 19570429,00.html

Time Magazine (1983, Mar 7). *Pentagon maverick Franklin Spinney* [Cover illustration]. Retrieved January 27, 2010, from http://www.time.com/time/covers/0,16641, 19830307,00.html

Tragesser, J. N. (1985, Jan-Feb). MANPRINT: manpower and personnel integration. *Army Research, Development, and Acquisition Magazine*, 4−6.

U.S. Army Center of Military History (2008). *Defense acquisition history project working paper #3: the evolution of DoD Directive 5000.1 acquisition management policy 1971 to 2003*. Retrieved April 8, 2010, from the Defense Acquisition History Project Web site: http://www.history.army.mil/acquisition/research/working3.html

U.S. Army Materiel Development and Readiness Command (1979). *Engineering design handbook, Army weapon systems analysis, part 2* (DARCOM Pamphlet 706-102). Alexandria, VA: Author.

U.S. Army Science Board (1981). *Army science board 1981 summer study, equipping the Army 1990–2000* (Vol. 1). Washington, DC: Author. (DTIC Accession No. ADC026511)

U.S. Army Science Board (1981). *Army science board 1981 summer study, equipping the Army 1990–2000* (Vol. 2). Washington, DC: Author. (DTIC Accession No. ADC026973)

U.S. Army Science Board (1982). *Army science board 1982 summer study, science and engineering personnel*. Washington, DC: Author. (DTIC Accession No. ADA125780)

U.S. Army Science Board (1983). *Army science board 1983 summer study, the future development goal*. Washington, DC: Author. (DTIC Accession No. ADA135819)

U.S. Army Science Board (1984). *Army science board 1984 summer study, leading and manning Army 21*. Washington, DC: Author. (DTIC Accession No. ADA150444)

U.S. Congress (1986). Independent cost estimates; operational manpower requirements, *Title 10, United States Code, Section 2434*. Washington DC: Government Printing Office.

Watts, B. D. (1996). *Doctrine, technology, and war*. Maxwell Air Force Base, AL: Air War College. (DTIC Accession No. ADA331985)

Weisz, J. D. (1967). *Systems analysis: human factors research and development contributions* (Technical Note 6-67). Aberdeen Proving Ground, MD: Human Engineering Laboratory. (DTIC Accession No. AD818566)

Weisz, J. D. (1968). *Systems analysis: manpower resources/system design integration methodology* (Technical Note 9-68). Aberdeen Proving Ground, MD: Human Engineering Laboratory. (DTIC Accession No. AD675481)

Weisz, J. D. (1986). MANPRINT: a historical perspective. *MANPRINT Bulletin*, 2, 3−4.

Weisz, J. D. (1989). *Human Engineering Laboratory support of MANPRINT* (Technical Note 1-89). Aberdeen Proving Ground, MD: Human Engineering Laboratory. (DTIC Accession No. ADA205830)

Wild, J. W. (1987). *The Army of Excellence: How ready?* (Student essay). Carlisle Barracks, PA: U.S. Army War College. (DTIC Accession No. ADA182850).

THIS PAGE INTENTIONALLY LEFT BLANK

IV. THE HUMAN SYSTEMS INTEGRATION TRADE SPACE PROBLEM

> The alternative, we believe, is to offer not just limits or constraints, but "trade offs." Indeed, we suggest that the success of the field of human factors will be proportional to the ability of the profession to provide such trade offs (Kennedy, Jones, & Baltzley, 1989, p. 1).

A. INTRODUCTION

Since the domains of human systems integration (HSI) are interrelated, any action affecting a single domain will necessarily propagate to one or more other domains, causing either desired or unintended effects. To help illustrate this idea, let us consider the analogue of a simple physical system such as the pulley system depicted in Figure IV-1. Le Chatelier's Principle asserts that if a set of forces are in equilibrium and a new force is introduced then, in so far as they are able, the existing forces will rearrange themselves so as to oppose the new force (Eigen & Winkler, 1981). The left panel in Figure IV-1 depicts three forces that are in equilibrium. Moving to the right panel, a fourth force is introduced and the original three readjust to a new point of equilibrium for all four. While this principle is unexceptional for physical systems, Hitchins (1992) contends that "the principle applies equally to interaction between economic, political, ecological, biological, stellar, particle or any other aggregations which satisfy the definition, *system* [emphasis in original]" (p. 61).

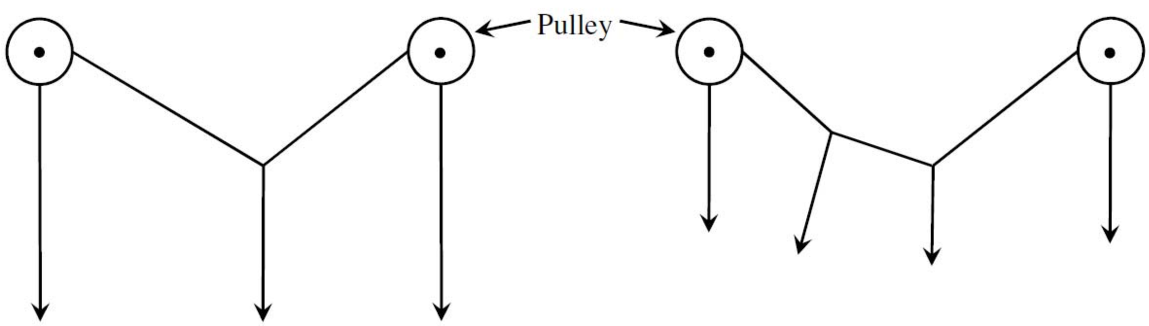


Figure IV-1. Forces in equilibrium in simple pulley system [From Hitchins, 1992].

We can extend the metaphor of the physical system in Figure IV-1 to contemplate how changes in one HSI domain might be resisted by corresponding changes in other HSI domains. An example of interacting HSI domains seeking a new equilibrium was given earlier in Chapter I (Figure I-2) based on the work by Miller and Firehammer (2007). Decreasing manpower (manpower domain) onboard U.S. Navy ships to lower life-cycle costs can result in short-term advantage. However, this is replaced by long-term economic loss as increased workload and fatigue (survivability domain) drive lower productivity (new system equilibrium) and more mishaps (system safety domain).

Another conceptualization of interacting HSI domains is illustrated by the vector diagram in Figure IV-2. In the left panel is a set of interacting HSI domains depicted as individual vectors, supposedly in equilibrium, such that the overall performance vector is as shown by the heavy arrow. In the right panel, a putative environmental disturbance or new system constraint is introduced that has the potential for changing the status quo. In so doing, it will perturb the equilibrium in an undesirable way. This unwanted perturbation may be managed by introducing complementary HSI domain changes, as shown, which have the net effect of cancelling out the unwanted effect(s). Such cancelling may be complete or simply sufficient to enable control of the interacting set of HSI domains as they tend towards a new point of equilibrium.

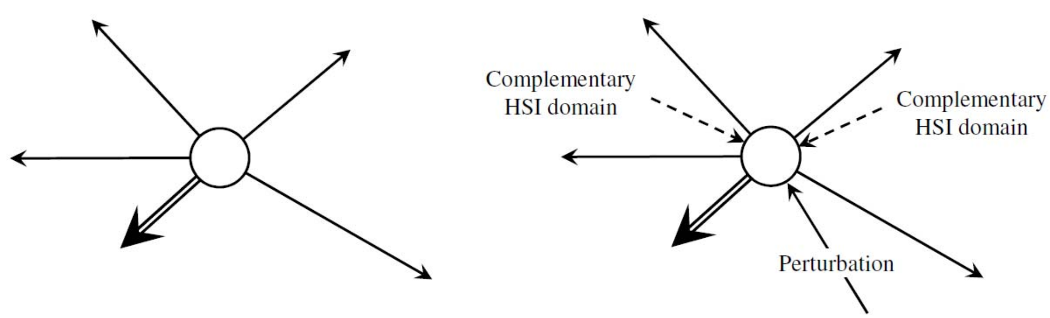


Figure IV-2. Complementary HSI domain inputs neutralizing unwanted perturbations [After Hitchins, 1992].

Overall, this notion of perturbations and choice of complementary HSI domains illustrates the existence of a trade space. This reality then necessitates a holistic

perspective of the performance trade space formed by the synthesis of the HSI domains, and as a consequence, the consideration of individual domain interventions in terms of tradeoff decisions. Such a worldview, or *Weltanschauung*, is central to the Naval Postgraduate School's definition of HSI, which calls for making "explicit the underlying *tradeoffs* [emphasis added] across the HSI domains, facilitating optimization of total system performance." However, current HSI manuals and handbooks do not provide much guidance on HSI tradeoffs; nor is there a well-established body of knowledge addressing HSI domain tradeoffs despite the obvious need (Barnes & Beevis, 2003). For example, Booher (1990, pp. 12, 42) makes only two references to "tradeoff," and the National Academies, through its Committee of Human Factors' HSI report (Pew & Mavor, 2007, pp. 3, 19, 34, 140), makes only four references to "tradeoff," none of which even begin to scratch the surface of the issue.

B. THE NAVAL POSTGRADUATE SCHOOL HSI PROCESS MODEL

1. Traditional Conceptual Models of HSI

Models are designed to represent a system under study, serve as an idealized example of reality, or explain essential relationships (Blanchard & Fabrycky, 2006). Thus, it should come as no surprise when Booher (2003) observes that HSI is often described in terms of "a top level conceptual model" (p. 4). One of the more common conceptual models of HSI focuses on the constituent human related disciplines that should be considered during system definition, development, and deployment. This type of HSI model can be traced back in some form or fashion to the Defense Department's human engineering community and generally takes the form of the model shown in Figure IV-3, which was taken from the U.S. Army's 1976 human factors engineering regulation (Army Regulation 602-1). This type of model is useful for conveying the scope of potential HSI considerations and ensuring adequate representation of relevant domain stakeholders. However, it also has the potential danger of leading those individuals who are predisposed to reductionist thinking to view HSI predominately as a collection of domain considerations to be addressed rather than as a complex, interacting system of domains to be managed. It is interesting, then, to observe that a later iteration

of the model, appearing in the Defense Department's 1991 acquisition instruction, makes inter-domain tradeoffs a more explicit part of the model (Figure IV-4). Nevertheless, despite its shortcomings, the historical longevity of this model suggests that it has utility as a top-level concept for HSI, particularly with regards to obtaining representation of all human-related disciplines, and consequently, we should be in no rush to discard it.

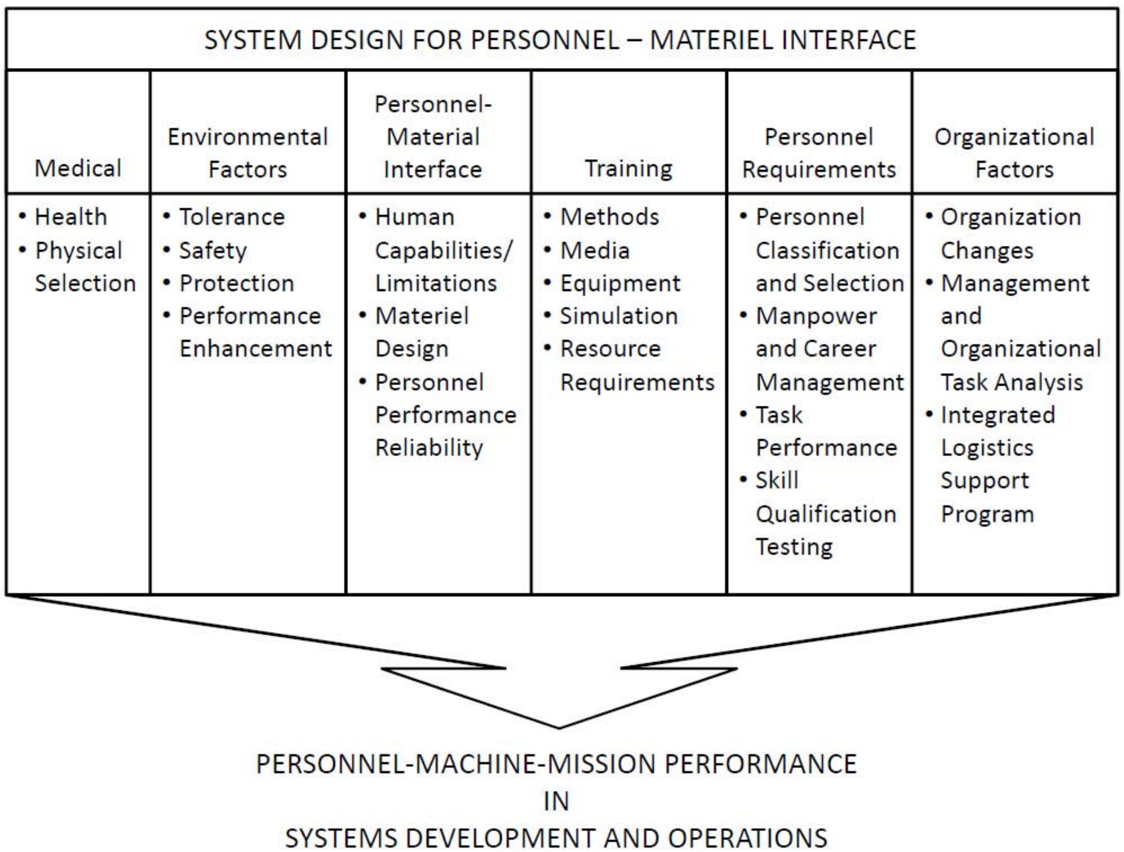


Figure IV-3. Early conceptual HSI model focusing on domain representation [From Department of the Army, 1976].

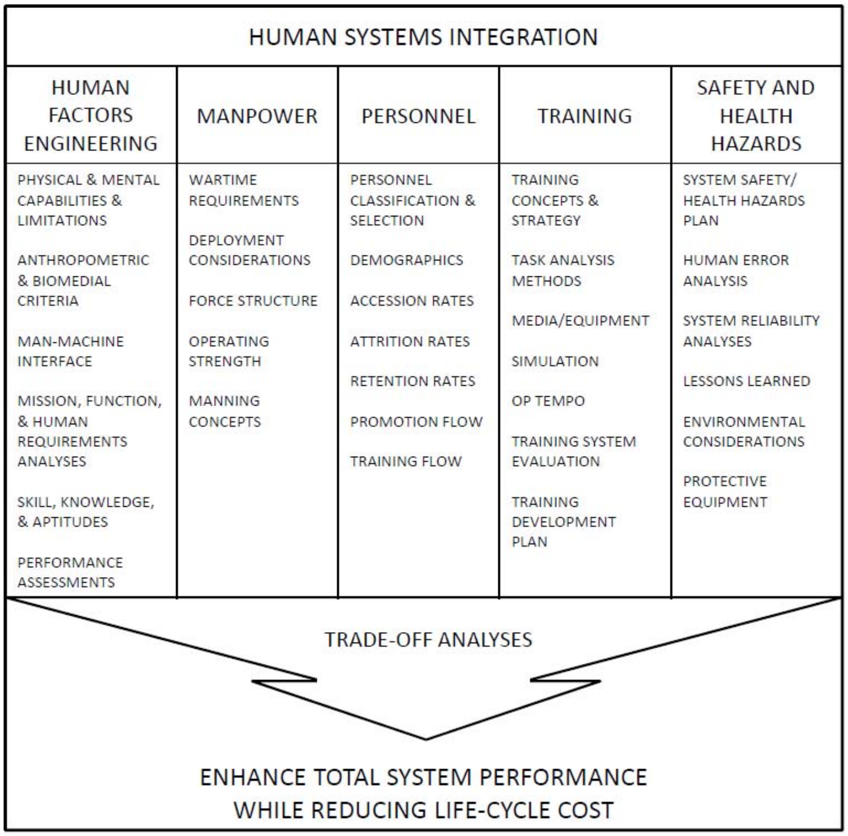


Figure IV-4. Conceptual HSI model focusing on domain representation and the need to consider inter-domain tradeoffs [From Department of Defense, 1991].

Another model frequently used to explain the HSI concept is that proposed by Booher (2003) and revised by Booher, Beaton, and Greene (2009). Booher builds on the earlier human engineering derived HSI models with the aim of explaining how the HSI concept is fully compatible with those systems engineering processes relevant to systems definition, development, and deployment and their life-cycle phases (Figure IV-5a). According to Booher (2003), as a top-level model, HSI brings two novel features to the systems engineering model: (1) a highly concentrated user focus in all aspects of the systems definition, development, and deployment stages; and (2) the application of human-related technologies and the HSI disciplines throughout the systems engineering management and technical processes. No system, product, or equipment inputs can be considered as having had an adequate consideration of the human component if it does not pass through the HSI process modulated with these two inputs. Booher

conceptualizes the HSI process as a double-integration process where both integration steps are modulated by human-related technologies and disciplines and driven by user requirements (Figure IV-5b). The first integration step creates a common focus for all seven HSI domains. At the second integration step, HSI contributions are fully integrated within the decision processes utilized by systems engineering and management processes.

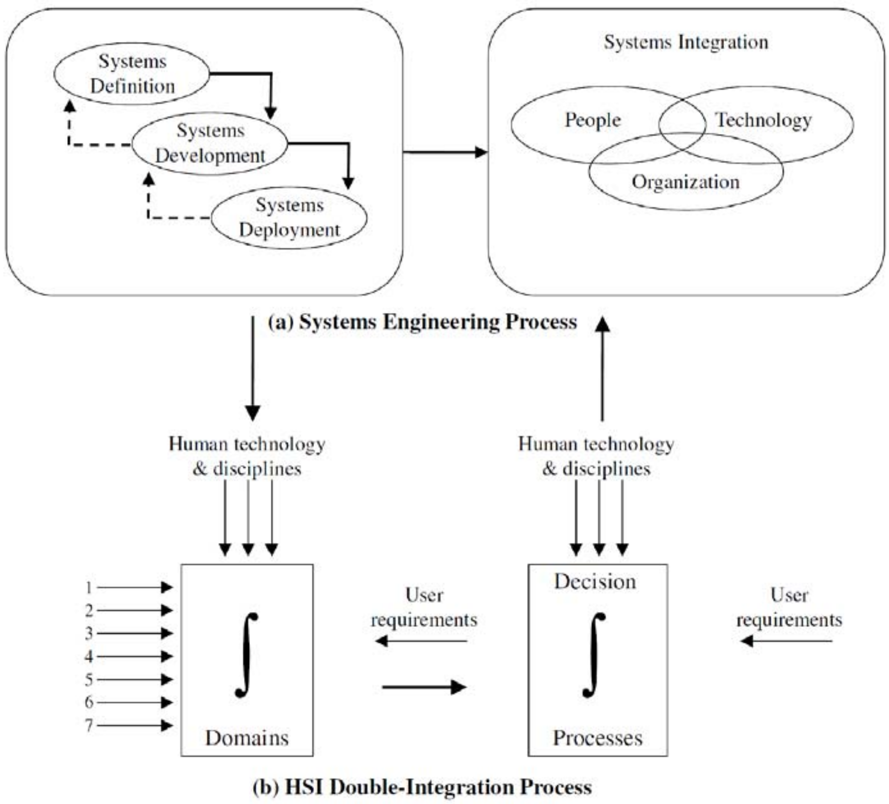


Figure IV-5. Systems engineering and HSI processes [From Booher, Beaton, & Greene, 2009].

Booher’s model is an evolutionary advance over the constituent domain-focused model in that it lends itself more to a synthetic view of HSI: a move from Descartes’ reductionism and Machine Age thinking to expansionist Systems Age thinking. With that said, since system design proceeds iteratively through a cycle of analysis, evaluation, and synthesis (Blanchard & Fabrycky, 2006), it may be more appropriate to view the two HSI models as complementary rather than as a parent-progeny couplet. Still, it appears to be a struggle for many novice HSI practitioners, and certainly for the majority of

students in the Naval Postgraduate School's HSI degree program, to conceptualize the activities that are occurring in the integration processes. For many students, the challenge is to reconcile the complexity of detail in the domain-focused model with the simplified abstraction of the Booher model. Consequently, this led the faculty and students at the Naval Postgraduate School to begin working on potentially better ways to illustrate the HSI trade space, in turn trading off complexity of detail for insight and understanding.

2. A Process Model of HSI

The emphasis in the Naval Postgraduate School's approach to conceptualizing HSI is to focus on the integration of the domains and the ability to leverage interactions between domains rather than pursue the additive value of each domain independent of the others. This approach implicitly recognizes the interdependency between domains and acknowledges the necessity for maintaining a holistic perspective of the potential solution space. In formulating their approach, the Naval Postgraduate School faculty looked to the field of biology for natural metaphors on which to build by analogy—an approach that is entirely congruent with the thinking of sociotechnical system theorists, who tend to view organizations as open and living systems, much like a biological cell (Pasmore & Sherwood, 1978). In this case, the metaphor involves an ecosystem of interdependent organisms. Bar-Yam (2003) suggests that such networks of dependency as are seen in ecosystems are generally characteristic of complex systems. Consequently, the existence of such networks of dependency requires that one think about patterns of emergent behavior in addition to the behaviors of individual network components—and thus the basis for the metaphor.

What follows is a description of a marine ecosystem as originally presented by Miller and Shattuck (2007). In a marine ecosystem, a multitude of natural and man-made conditions, such as storms, pollution, and overfishing, stress the delicate balance required for a healthy ecosystem. If one is interested in the health of the ecosystem, one could examine a number of metrics reflective of the physical and chemical composition of the ecosystem: water temperature; concentrations of nutrients, chemicals, and

pollutants; oxygenation; etc. However, there are other indications of the state of an ecosystem: scientists look for the presence of “indicator species” when trying to characterize the integrity of a marine ecosystem. For example, the presence of large numbers of marine mammals, such as dolphin, is indicative that the ecosystem is healthy since it is able to sustain a population of marine mammals with voracious appetites. Absence of such indicator species may indicate problems with the balance of the ecosystem since these indicator species will only be present when there is an adequate food supply. Thus, disruptions in the integrity of the marine ecosystem can be inferred by observing the movement patterns of such indicator species.

The relationships in a marine ecosystem are represented in Figure IV-6. The rectangle simply delineates the arbitrary boundary of our system-of-interest. Our system is clearly an open system so this is a very permeable boundary at best. Within our system-of-interest, only certain components and dependencies are identified by the text and arrows respectively. The components and dependencies shown are not meant to be all-inclusive. Obviously, a marine ecosystem is a highly complex system with many more components, which in turn, have many interactions and dependencies. Thus, we have merely used the technique of simplification and abstraction to make the complexity manageable—exactly the same technique that is used by systems engineers (Bar-Yam, 2003). The important observation is that we can make reasonable inferences about the state of such a complex system (i.e., ecosystem integrity) based on an assessment of intermediate level emergent effects (e.g., presence of indicator species). Additionally, when intermediate level emergent effects are undesirable (e.g., there are no dolphin), then we need to look back through the network of dependencies to find the root cause(s). Moreover, we need to appreciate the underlying network of dependencies if we are to have any chance of successfully predicting the effects of changes to system inputs or constraints.

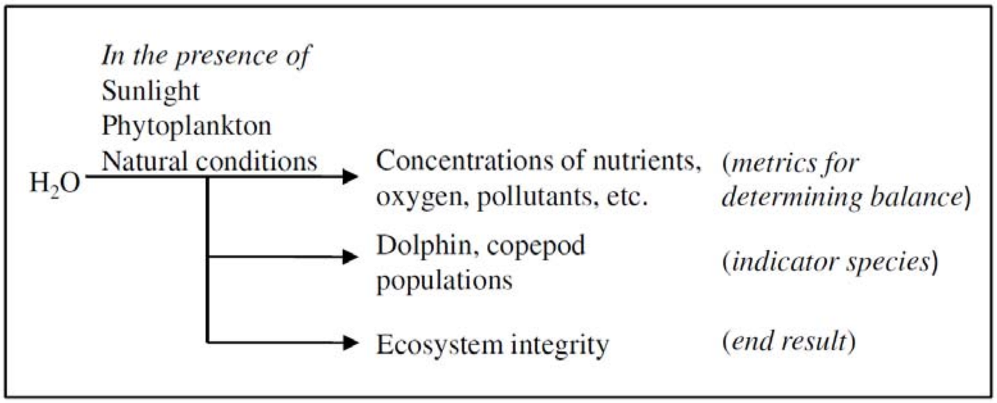


Figure IV-6. Simplified model of a marine ecosystem [From Miller &Shattuck, 2007].

This ecosystem metaphor is offered mainly to assist in the understanding of the proposed HSI process model that is represented in Figure IV-7. As described by Miller and Shattuck (unpublished manuscript):

> …this new approach to HSI assumes that some domains are inputs into the system while others are products or by-products that naturally result from decisions made regarding the inputs. The tradeoff decisions made in any part of the process will result in dramatic differences in the solution space or trade space. Adopting this new approach to HSI yields another outcome in that is forces the decision maker to be explicit about the tradeoffs that are being made in the HSI process. This transparency in tradeoffs is a critical part of the new model and implementation of HSI (p. 9).

Figure IV-7 depicts the entire system-of-interest when viewed as an open system comprised of HSI domains and emergent system effects. Again, the large rectangle represents the open system boundaries, delineating the system-of-interest as determined by the system definition to include system requirements and capabilities. The mission(s) and concept of operation(s) in which the system is to function are also part of the context within which HSI tradeoffs occur.

On the left side of the model are the domain considerations that are inputs to the system-of-interest: task design (i.e., human factors engineering), manpower, personnel, and training. These elements describe the personnel subsystem and its interface with the technological subsystem that together comprise our system-of-interest. According to

sociotechnical systems theory, the personnel and technological subsystems should be jointly optimized so as to exhibit the emergent properties that make the system-of-interest of value to its stakeholders. This is a constrained optimization at best, requiring the constant management of the tensions between performance, cost, schedule, and risk. These tensions are analogous to the storms, pollution, and overfishing that stress the delicate balance required for a healthy ecosystem.

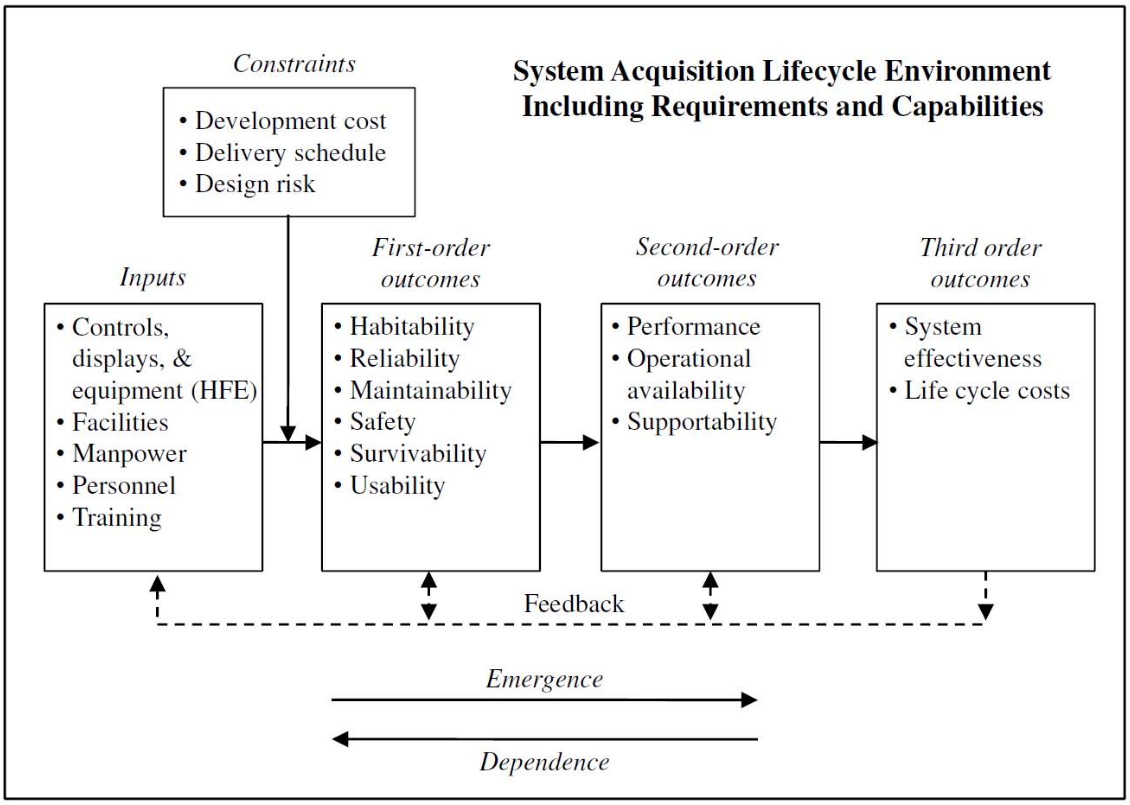


Figure IV-7. The Naval Postgraduate School HSI process model with modification by the author [After Miller & Shattuck, 2007].

Moving to the right in the HSI process model, the first-order outcomes are the emergent properties that result from the attributes and relationships of the inputs. Many of these outcomes are adjectives, describing desired qualities. For example, close coupling of the personnel, training, and human factors engineering domains will result in interfaces between the personnel and technological subsystems that exhibit good usability. It is important to note that several HSI domains, those being habitability,

safety, and survivability, are emergent properties of the joint optimization of the personnel and technological subsystems. As a result, these domains are necessarily dependent on the state of the input domains, and hence, the illogic of trying to consider them as independent entities. These first-order domains cannot be purchased or procured directly as an input or consumed as a resource. As a thought experiment, how would one purchase a mishap rate of less than 0.0001 per thousand hours of operation? Inevitably the answer brings one back to (one or more of) the input domains. Consequently, we can use these emergent properties to assess the quality, or health, of the joint optimization of the personnel and technological subsystems. Just as we used metrics such as nutrient and oxygen concentrations to evaluate the health of a marine ecosystem, we can monitor metrics like noise, temperature and humidity to assess the health of a system in terms of habitability. Continuing the analogy, safety can be thought of as an indicator species: its presence is indicative of the ability to sustain reliable work, while its absence suggests an imbalance within or between the input domains.

As we continue further right in the HSI process model, we move up layers in the hierarchy of complexity and observe increasingly more holistic measures of system health. Admittedly, the model omits detailed consideration of the hardware and software inputs to the technological subsystem at large, but this is in keeping with Booher's highly concentrated user focus. One should also note that higher order outcomes become progressively less user-centric and more reflective of total system performance, thereby showing the dependence of total system performance on the quality of the human-centric input domains.

The HSI process model builds on the two traditional conceptual models of HSI that were discussed in Section IV-B1. Miller and Shattuck assert that the main contribution of the HSI process model is that it gets into the details of how both domains interact and tradeoffs might be exercised:

> Decisions made in any one of the HSI domains—or in cost, schedule, and risk—will propagate throughout the system and will impact the other domains, creating a sort of domino effect whereby one decision may affect something that is seemingly far-removed from the original decision. Ultimately, all these decisions will affect total system performance

> although this effect is not always obvious. The ability to visualize and understand the ramifications and consequences of decisions, and the inherent tradeoffs that occur throughout the acquisition process, is the very essence of HSI (p. 12).

It remains an area of active research at the Naval Postgraduate School to develop practical applications of this process model. Nevertheless, it is still not evident how any of the HSI conceptual models discussed so far could address the local versus global—or micro versus macroergonomic—duality of the HSI trade space that was discussed in Chapter II. Thus, we next turn to work by DePuy and Bonder addressing this very duality from the more circumscribed context of manpower, personnel, and training.

C. HSI AS THE PROCESS OF MANAGING THE SUPPLY OF AND DEMAND FOR MANPOWER, PERSONNEL, AND TRAINING (MPT) RESOURCES

1. MPT Supply and Demand Management

Describing the Army's acquisition of new systems and equipment in the early 1980s, DePuy and Bonder (1982) stated:

> The new systems, more often than not, involve high technology in an effort to achieve high performance -- performance which provides an advantage over the postulated enemy threat and adversary battlefield systems. […] With very few exceptions personnel requirements derived from these new systems require higher aptitudes and more training for operators and especially for maintainers…A slow but steady "skill creep" is very much in evidence (p. 2).

It is likely that many people within the Defense Department would view this perspective as still being valid today. The current emphasis on minimizing manpower necessitates job consolidation, thereby increasing the breadth and depth of tasks assigned to individuals. DePuy and Bonder assert that it is frequently left to the MPT communities to address this skill creep. This pattern occurs because the supply side of the problem (i.e., managing human resources), although difficult, is more immediately tractable than the demand side (i.e., requirements). They caution, however, that this is not a sustainable approach in the long term:

> Personnel managers can turn to recruiters and ask for the accession of higher quality personnel. They can turn to the trainers and ask for increased skill performance through improved or longer training. Although both actions are necessary, neither is apt to suffice in the long run. In the long run it will probably be necessary to reduce skill demand by intervention in the weapons system acquisition process (pp. 3–4).

The crux of the problem, as DePuy and Bonder see it, is a failure to effectively match skill requirements (demand) with skill availability (supply). What is then needed is a system for "closing the loop between the MPT planning and analysis process and the system design process (the design engineers)" (p. 10).

Taking an operations research perspective, DePuy and Bonder frame the problem in terms of MPT supply and demand management. Figure IV-8 depicts a concept map of their description of *MPT supply management*.

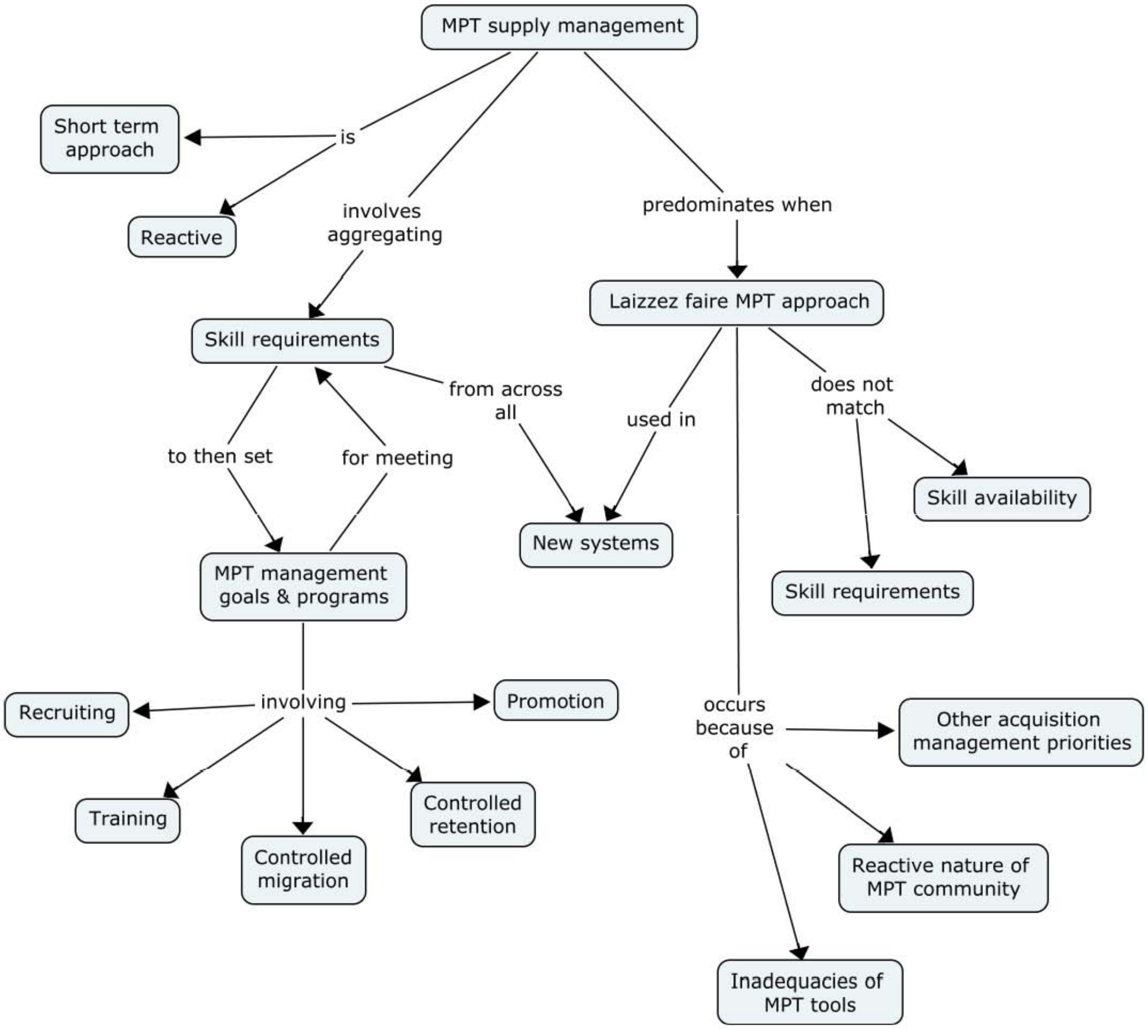


Figure IV-8. Concept map for MPT supply management as described by DePuy and Bonder (1982).

The supply-side approach is a reactive mode that consists primarily of efforts to forecast MPT requirements so that the recruiting and training organizations can be focused on meeting skill requirements. It occurs when a *laissez faire* approach is taken to MPT requirements, such that program managers and design engineers are forced into taking the lead in addressing MPT issues. While an individual program may do an exemplary job in addressing MPT issues, the result over the aggregate of programs is still an undisciplined front-end process. When faced with the very real difficulties of *concurrently* addressing MPT issues early in the system development process, program

management are apt to follow a *sequential* approach that is conceptually simple and inherently less risky to themselves—but which, in the words of DePuy and Bonder, is "a cumulative disaster" for a military service at large. In this approach, potentially superior technologies are first chosen, advanced engineering models are then built, task/skill analyses and limited human-machine tradeoffs are performed with the primary objective of keeping operator and maintainer tasks under control, technical manuals are developed, training is addressed, and finally any residual MPT issues are finessed during initial system deployment. This approach, when applied to system after system, has been shown to lead to the rapid accumulation of MPT demands that can surpass the capability of a military to satisfy.

The alternative, proactive approach is *MPT demand management*, an overview of which is provided by the concept map in Figure IV-9. The simple logic underlying demand management dictates that future MPT constraints serve as shaping factors in the selection of system concepts by system developers and force structure designs and modernization schedules by force developers.

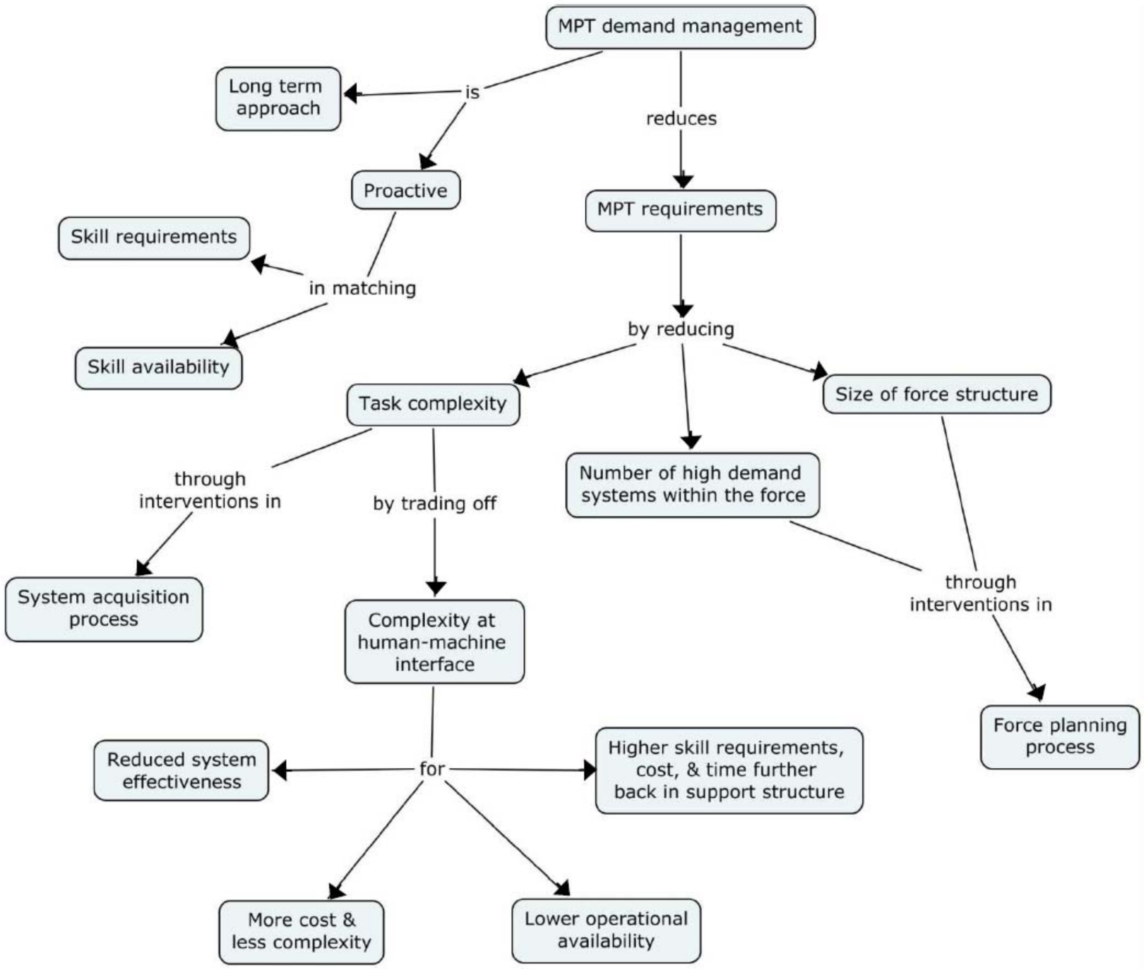


Figure IV-9. Concept map for MPT demand management as described by DePuy and Bonder (1982).

From the standpoint of demand management, there are three principle ways to reduce or manage MPT requirements:

1) Engineer complexity out of the human-machine interface, thereby reducing or simplifying the operator and maintainer tasks performed by service members
2) Reduce the number of high demand systems within the force—that is, changing the internal composition of the force
3) Reduce the size of the force structure and carry the demand downward across all systems.

The latter two approaches are strategic management decisions. However, the first approach, which can be addressed through the human factors engineering domain, is no trivial or simple matter. System developers looking to reduce complexity in the human-machine interface are likely faced with a zero-sum game. They can trade off complexity at the human-machine interface for some combination of the following: reduced system effectiveness; more cost and less complexity; lower operational availability; and higher skill requirements, costs, and time expended elsewhere in the system support structure. Thus, it is critical that those on the supply side of the problem ensure that MPT demands inherent in higher performance systems cannot be met through supply-side interventions before inducing one or more of the unattractive tradeoffs available to system developers. However, individual program MPT decisions must also be considered from the perspective of aggregate MPT supply and demand. As DePuy and Bonder aptly note, “Of course the Army can meet the demands for any one system, but the problem is to meet the aggregate demand of all systems” (p. 8).

2. A Model for Interfacing MPT Supply and Demand

DePuy and Bonder’s prescription to skill creep was a dynamic process for managing the interactions between MPT demand (requirements) and MPT supply. They described their process in terms of a conceptual model for interfacing MPT supply and demand, both at the *macro* multi-system level and at the *micro* system-specific level, within the context of the acquisition process. In so doing, their objective was to move beyond the acquisition system’s historical focus on performance, cost, and schedule. They believed it was absolutely necessary to also consider, in an organized and repeatable manner, whether sufficient MPT resources would be available to operate and maintain future systems to their intended design performance.

a. The Macro MPT Supply and Demand Interface Process

Figure IV-10 is a schematic representation of the interface between MPT supply and demand at the aggregated level of a military service. The left side of the figure depicts the MPT demand process, the right side the supply process, and the center their interface. MPT demand is generated both by the characteristics of new systems and

the number of those systems that will be integrated into the force structure. As depicted in the upper left of Figure IV-10, the SYSTEM ACQUISITION PROCESS employs various studies (e.g., mission area analyses, trade off analyses, best technical approaches, etc.) and documents (e.g., initial capabilities documents, capabilities development documents, etc.) to move new systems (e.g., SYS A, B,…,X) through the development phases (e.g., Material Solution Analysis, Technology Development, etc.) and on to the FORCE DESIGN PROCESS where they are integrated into the total force. The design characteristics of these new systems determine the allocation of functions between the human and machine, and hence, drive the personnel and training requirements for the systems.

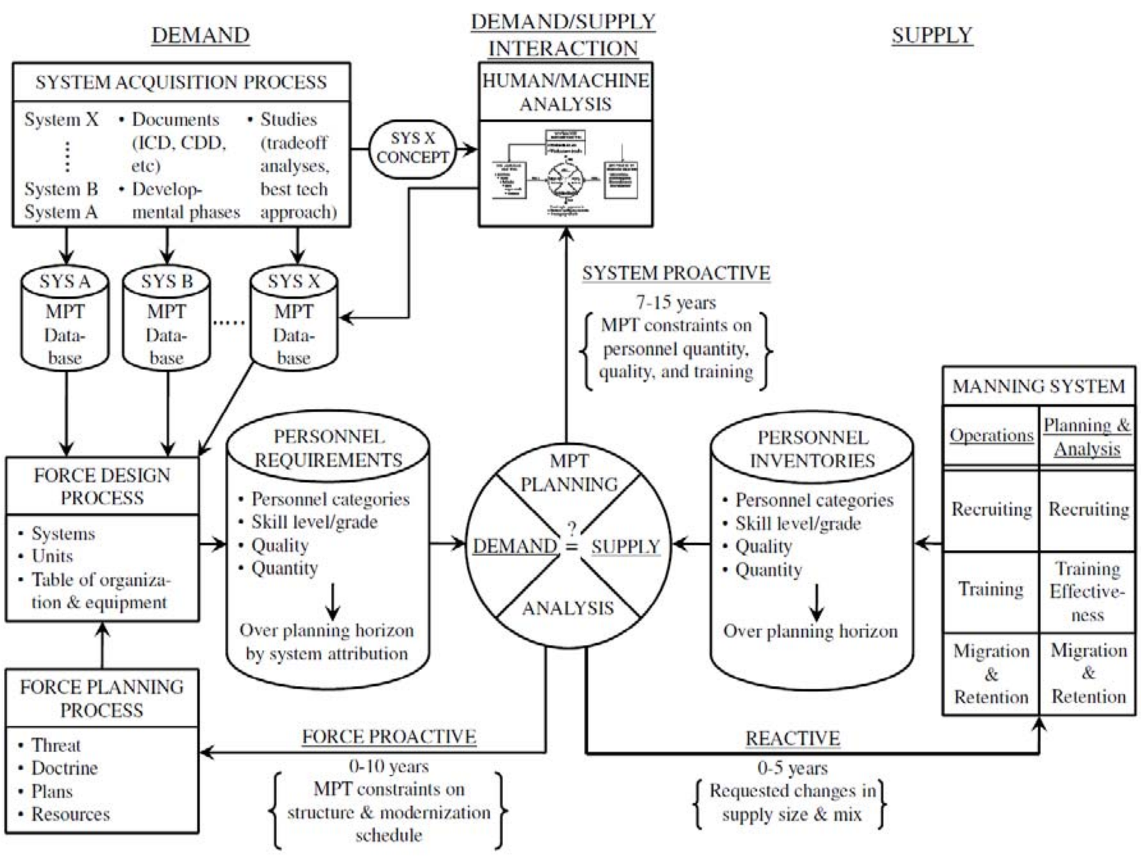


Figure IV-10. Schematic representation of the interface between MPT supply and demand at the aggregated level of a military service [After DePuy & Bonder, 1982].

The size of the force and its functional configuration are derived from the FORCE PLANNING PROCESS, which relates force stength and structure to threats, doctrine, and available resources. Thus, the force planning process drives manpower requirements for each personnel category. The product of the FORCE DESIGN PROCESS is a time-phased personnel inventory by unit type and number and a table of organization and equipment for each unit, within which the the systems, their operators, and maintainers, and leadership are integrated. From a MPT perspective, the resulting demand signal is expressed in terms of a PERSONNEL REQUIREMENTS database containing the total, time-phased MPT requirements for a military service.

For a single system, MPT information generated by the SYSTEMS ACQUISITION PROCESS is stored and continually updated in a SYSTEM MPT DATABASE. The system MPT databases contain current, system-specific MPT requirements information such as number of personnel by category; functions, tasks, and subtasks to be performed by personnel; skill level requirements; training requirements; etc. Accordingly, the system MPT databases provide access to current MPT requirements as systems progress through the development process and push information through the FORCE DESIGN PROCESS to develop aggregated demand data.

Moving to the right side of Figure IV-10, the MPT supply process is embodied in the MANNING SYSTEM, the operation of which affects flows of personnel into a military service (i.e., accession), among various personnel categories (i.e., migration), and out of a service (i.e., separation). The number and type (e.g., demographic characteristics, personnel category, rank, skill level, etc.) of service members that are available is a function of recruiting, reclassification, migration, and retention incentives; the effectiveness of training systems; and the demographic characteristics (e.g., aptitude, education level, etc.) of those individuals to whom that training was applied. The product of the manning system is a MPT supply signal expressed in terms of a PERSONNEL INVENTORY database containing the time-phased inventory of trained personnel available for assignment.

The aggregated interface between MPT supply and demand is accomplished by the MPT PLANNING ANALYSIS (center of Figure IV-10), which

continually identifies time-phased personnel shortages by personnel category over a sufficiently broad planning horizon (i.e., 0−20 years). Given a set of MPT supply/demand imbalances, decision makers can attempt to alleviate the problem in a REACTIVE manner by making short-term (i.e., 0−5 years) supply side fixes. This would involve time-phased changes in the manning system; recruiting, reclassification, migration, and retention flows; and/or training programs to alter the size and/or mix of future personnel inventories. Alternatively or concurrently, decision makers could initiate PROACTIVE changes by making longer-term demand side fixes. FORCE PROACTIVE actions involve midterm (i.e., 0−10 year) MPT constraints on force structure size and composition and modernization schedules. For example, decision makers could stretch out time-phased system MPT requirements and the system development status for immature systems and/or change initial/full operational capability dates for mature systems to "smooth out" shortages. In the extreme, the overall force design could be adjusted by reducing the number of systems in the force. For systems in the very front end of the development cycle, SYSTEM PROACTIVE actions are pursued to head off projected future shortages. Such long-term (i.e., 7−15 years) MPT constraints typically include constraining the number and/or skill levels of personnel that can be designed into a system. While DePuy and Bonder discuss the use of these demand side changes primarily in terms of alleviating MPT shortages, there is no reason they could not also be used to drive force-shaping objectives or meet organizational investment goals.

b. The Micro MPT Supply and Demand Interface Process

In addition to the multi-system, service-wide aggregated interface between MPT supply and demand, integration of MPT into the systems acquisition process requires that supply and demand be interfaced at the system level early in its conceptual design. DePuy and Bonder suggest that this latter MPT supply/demand interface occurs in the HUMAN/MACHINE ANALYSIS process depicted schematically in Figure IV-11. Based on a systems engineering functional analysis, an initial HUMAN/MACHINE TRADEOFF ANALYSIS is performed to allocate functions between the human (e.g.,

operator, maintainer, supervisor, etc.) and machine components, thereby formulating an initial conceptual design (e.g., SYS X CONCEPT). This analysis should give consideration to various criteria derived from a stakeholder analysis such as desired system performance, costs, critical system attributes, etc. This initial conceptual design is then subjected to both a MPT FRONT END ANALYSIS and a MPT CAPABILITY TRADEOFF ANALYSIS.

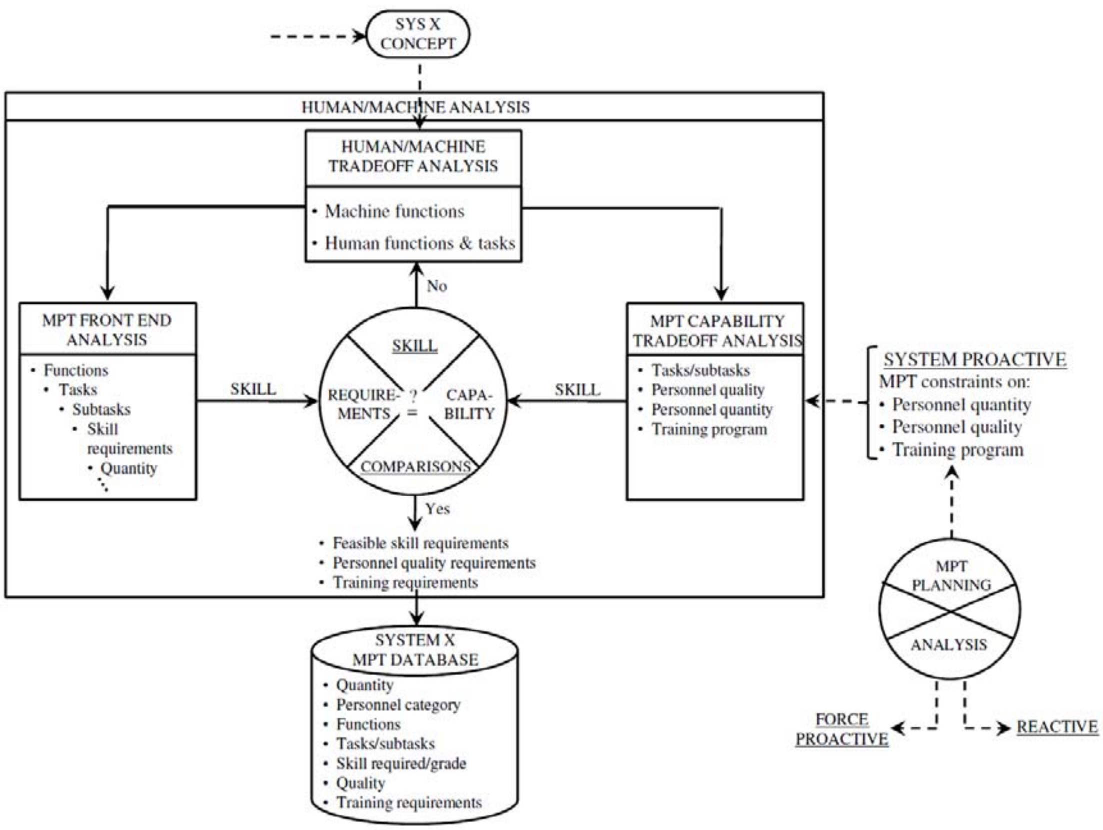


Figure IV-11. Schematic representation of the interface between MPT supply and demand at the system-specific level [After DePuy & Bonder, 1982]).

The purpose of the MPT front end analysis is to determine the specific human performance requirements deemed necessary to satisfactorily accomplish the functions and tasks allocated to the human in the human/machine tradeoff analysis. The definition of "satisfactorily accomplish" should be derived from the stakeholder analysis

and the resulting value model(s). The MPT front end analysis utilizes a hierarchical functional decomposition to identify the tasks, subtasks, and skill levels required to achieve the desired system performance. Neither the human/machine tradeoff analysis nor the MPT front end analysis should explicitly consider personnel quality or training issues. The procedures should focus on determining the human performance requirements by task/subtask. Thus, the product of the MPT front end analysis is a MPT demand signal in terms of the required capability for skilled human work.

The purpose of the MPT capability tradeoff analysis is to determine the human performance capability that can be provided to man the system. As envisioned by DePuy and Bonder, for each task or subtask allocated to the human, the MPT capability tradeoff analysis determines achievable levels of performance by examining tradeoffs between the number of personnel assigned the task or subtask, the quality of the personnel (as measured on some aptitude scale like the Armed Forces Qualifications Test (AFQT) or Armed Services Vocational Aptitude Battery (ASVAB)), and the training provided (e.g., type, duration, frequency, etc.) within the MPT constraints derived from the multi-system MPT PLANNING ANALYSIS shown to the right in Figure IV-11. Thus, the human-machine tradeoff analysis is embedded within, and constrained by, the larger macro MPT supply/demand interface (as shown center top in Figure IV-10)

The micro analog of the macro MPT planning analysis is the SKILL COMPARISON ANALYSIS (shown in the center of Figure IV-11), which compares the outputs from the MPT front end analysis (i.e., demand) and MPT capability tradeoff analysis (i.e., supply) by task and performance level to determine if the required human performance capability can be provided. If the skill comparison analysis indicates that there is a supply/demand imbalance, then the conceptual design process must be reiterated and the following possible tradeoffs considered:

1) Reallocating more tasks/subtasks to the machine component, possibly resulting in increased cost
2) Reducing the human performance requirements to a feasible level, resulting in decreased overall system performance

3) Relaxing the personnel quality requirements and/or training constraints to improve available human performance capability, but this will likely cause a tightening of MPT constraints on other systems because MPT resources are finite.

This conceptual MPT design process, involving the human/machine tradeoff analysis, MPT front end analysis, and MPT capability tradeoff analysis subject to proactive MPT constraints, should be iterated early in the acquisition process until the system's MPT requirements are consistent with forecasts of available personnel capabilities. Output from the dynamic interplay of MPT supply and demand is a feasible set of human performance requirements from the system design (MPT front end analysis), personnel quality requirements (MPT capability tradeoff analysis), and training requirements (MPT capability tradeoff analysis). All information generated in the human/machine analysis process is stored and updated in the corresponding SYSTEM MPT DATABASE, which in turn should provide updates to the PERSONNEL REQUIREMENTS database.

c. The MPT Capability Tradeoff Analysis

The purpose of the MPT capability tradeoff analysis is to provide information regarding available human performance capabilities. Since it is intended that the output of the capability tradeoff analysis be compared with that of the MPT front end analysis, personnel capabilities should be measured in terms of the performance levels that can be achieved on a task/subtask basis. As described by DePuy and Bonder, the MPT capability tradeoff analysis "procedures should facilitate examination of the tradeoffs among quantity of personnel, quality of personnel, and training level to achieve task/subtask skill levels, within constraints imposed on the quantity, quality, training [sic] dimensions by the multi-system supply and demand planning process" (p. A-9). Figure IV-12 depicts DePuy and Bonder's conceptual view of the types of information conveyed in a MPT capability tradeoff analysis. The quality dimension on the ordinate in Figure IV-12 should reflect existing personnel measures such as AFQT category, ASVAB score, etc. DePuy and Bonder assert that "the [capability tradeoff analysis] is a critical component in determining MPT requirements and in providing usable MPT guidance to

design engineers" (p. A-9). Thus, capability tradeoff analysis data goes a long way towards enabling the military services to adequately specify MPT requirements and evaluate contractor responses.

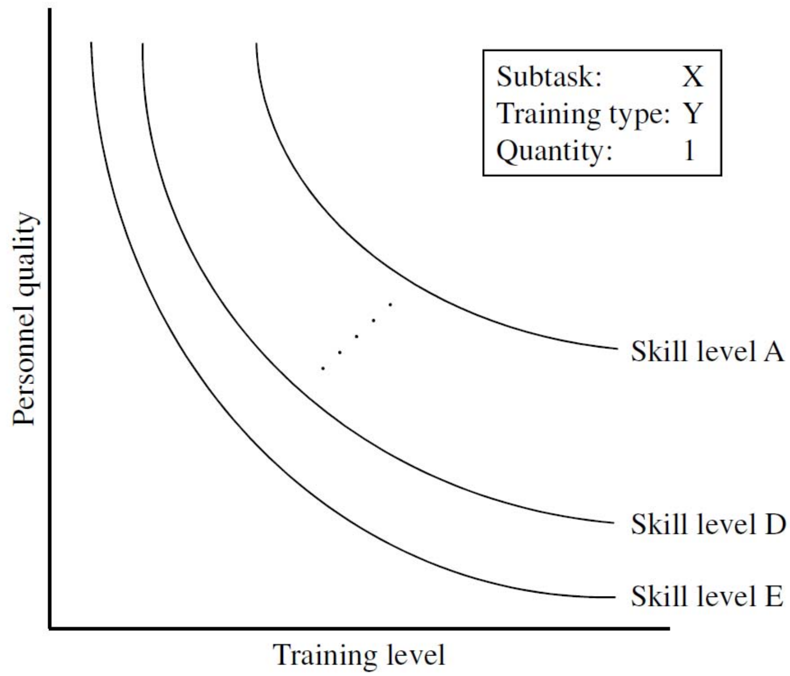


Figure IV-12. Achievable skill levels for subtask X as a function of personnel quality and training level [From DePuy & Bonder, 1982].

D. ISOPERFORMANCE

1. Considering Aptitude and Training Tradeoffs

Developing capability tradeoff analyses (i.e., tracing equivalent combinations of aptitude and training that yield a specified level of performance) is a highly applied problem. It would help a great deal to have known empirical regularities on which one could depend. It would also help to have a good theory of human performance that could support the development of analytical models relating aptitude, training, and performance. However, neither is necessarily sufficient to formulate a capability tradeoff analysis for a specific function or task. To serve the purposes of performing capability tradeoff analyses, it is not enough that an empirical regularity, such as the power law of practice, will hold on the average, for it must hold for the particular function or task of

interest. Further, it must be possible, given both the task and the regularity, to say just how well an individual will perform after a certain length or type of training. Theories of human performance must meet these same requirements. Unfortunately, at present, neither empirical regularities or theory meet these requirements. Therefore, in practice, system developers must hypothesize about human performance-related tradeoffs and then they should carry out experiments or tests to assess the veracity of their hypotheses—although the latter is not always done.

This conclusion does not mean that empirical regularities and sound theory are not helpful in performing capability tradeoff analyses. For example, the power law of practice provides us potentially useful insights into the personnel and training factors (determinants) postulated to contribute to task performance. According to this empirical regularity, skill acquisition tends to follow a relatively ubiquitous negatively accelerated learning curve, which can be represented by the following mathematical expression (Newell & Rosenbloom, 1981):

$$P = a + b \cdot \left(N + E \right)^{-r} \tag{1}$$

where P is the time taken to perform a task, a is the asymptote or highest level of performance obtainable, b is the performance on the first trial, N is the amount of practice in terms of trials, E is the transfer from prior experience or learning required to attain entry level performance, and r is a learning rate parameter (Table IV-1). Since practice is the canonical form of training, HSI training domain considerations are directly reflected in the N variable in the power law of practice. Likewise, training domain considerations of methods, curricula, and training system design will impact training effectiveness, and hence, the rate of improvement as reflected by the value of the r parameter. With regards to the HSI personnel domain, to the extent that an individual's experience with prior tasks is similar to the target task, positive transfer occurs (Wickens & Hollands, 2000) as captured by the variable E. Also, since aptitude tests predict proficiency on various tasks and propensity for a variety of types of learning (Matthews, Davies, Westerman, & Stammers, 2000), individual aptitudes will influence the values for the a, b, and r parameters in the power law. Overall then, the power law suggests both that the HSI personnel and training domains *must* interact and that their interactions may be complex.

Table IV-1. Variables in the power law of practice and corresponding HSI training and personnel domain considerations.

Power law		HSI domain considerations	
Variable	Description	Training	Personnel
a	Highest level of performance obtainable (asymptote)	--	Individual aptitudes
b	Performance on first trial	--	Individual aptitudes
E	Transfer from prior experience (in terms of equivalent number of practice trials)	--	Prior experience (knowledge, skills, etc.)
N	Practice in terms of number of trials	Quantity of training	--
P	Time taken to perform task	--	--
r	Learning rate parameter	Quality of training (training effectiveness)	Individual aptitudes

In terms of a theoretical perspective for considering personnel and training domain tradeoffs, aptitude-treatment interaction (ATI) theory provides a useful prototype. Briefly, the basic premise of ATI theory is that some instructional strategies, referred to as treatments, vary in their effectiveness for particular individuals depending upon the specific abilities of those individuals. Snow (1978) proposed that "individual differences in performance on ability tests and learning tasks are manifestations of cognitive processes to each" (p. 227). Snow's implication is that learning performance is higher when the learning method capitalizes on an individual's cognitive aptitudes. Likewise, achievement will be lower when the learning method requires that an individual use cognitive processes in which they are relatively weak or lacking (Whitener, 1989).

Aptitudes and treatments interact in two basic ways as depicted in Figure IV-13. The left panel shows the disordinal ATI, where one type of treatment yields high

achievement in individuals tending to one side of the aptitude spectrum, while another treatment helps individuals on the other side of the spectrum. The right panel shows the more common ordinal ATI, where one treatment is more effective overall, but it is particularly beneficial for individuals on one side of the aptitude spectrum (Whitener, 1989). Cronbach and Snow (1977) and Snow (1989), in reviewing the research on ATI, found that such interactions are very common in education. However, in practice, many aptitudes and instructional treatments interact in complex patterns that can be difficult to clearly demonstrate or understand. Although ATI research has focused predominately on acquisition of reading, language, mathematic, and science skills in the educational setting, it is plausible that the main conclusions should generalize to other skills (Matthews, Davies, Westerman, & Stammers, 2000). Nevertheless, for the system designer needing to perform a MPT capability tradeoff analysis for a particular function or task, such generalities about aptitude-treatment interactions will not suffice.

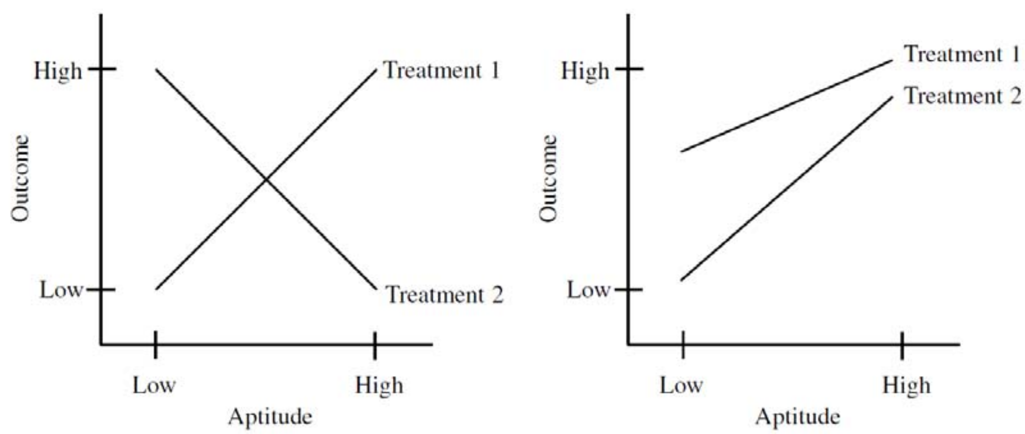


Figure IV-13. Possible interactive effects of aptitude and two instructional treatments on learning outcomes [From Whitener, 1989].

This situation is not unusual in the larger field of human factors engineering, and Meister (1999), in his history of human factors engineering, calls out this issue of quantitative prediction as one of the field's major unresolved problems. It often happens in applied work that there are no readily available empirical regularities or theories upon which one can rely for making real-world predictions. In such situations, if informed

decisions are then to be made, it can only be by studying empirically how performance varies as a function of its determinant. However, as noted by Jones, Kennedy, and Stanney (2004):

> If these determinants were few in number, not more than four, say, such functional relationships could be worked out by means of conventional designs (complete factorials) without great difficulty. Unfortunately, the determinants are rarely, if ever, few in number (p. 591).

This point has obvious implications for conducting capability tradeoff analyses. If we consider just the empirical regularity of the power law of practice, we have no less than six variables (Table IV-1), or determinants, that we need to assess if we are to make predictions about performance. How then to proceed? In the following section, I borrow heavily from a paper published by Jones and colleagues in *Presence* (Jones, Kennendy, & Stanney, 2004) to answer this question. While their paper focuses on cybersickness, the principles and methodologies discussed are generalizable to the larger issue of human performance.

2. Simon's Research Strategy

Charles W. Simon was among the first engineering psychology doctorates to flow from Paul Fitts' Laboratory of Aviation Psychology at Ohio State in 1952. Simon joined the RAND Corporation in Santa Monica and soon moved on to the Hughes Aircraft Company prior to joining the Canyon Research Group, where he worked with Jones and Kennedy on the Navy's Visual Technology Research Simulator project. Simon subscribed to the philosophy that the task of the engineering psychologist is to predict and control some performance of interest that occurs in the real world. He assumed that this performance of interest has many manipulable determinants of some importance, often more than five and likely at least 15–30. However, he also assumed that these determinants obey what he called the "Pareto maldistribution theory," meaning that the distribution of the magnitudes of their effects may be represented by a chi-square distribution. Thus, while performance may be affected by many determinants, only a few factors are critical.

Given then that some performance of interest likely has many manipulable determinants, the obvious first step is to screen those determinants to identify the most important ones. Simon's concern here is if only a few factors are studied at a time, the inevitable result is a list of factors with little evidence as to their relative importance or the interactions between them. Moreover, there can be no comprehensive accounting of how the performance of interest is functionally determined. Simon took the human factors community to task on this point, reviewing 239 analysis of variance tables published in the journal *Human Factors* during the period from 1958 to 1972. He found that the typical experiment investigated the effects of two factors, and less than 8% of the experiments studied four or more factors (Simon, 1976a). Over a period of 20 years, Simon published a series of technical reports, together totaling more than 1,500 pages, calling attention to the literature on advanced experimental designs and urging engineering psychologists to adopt them in their work (Simon, 1970, 1973, 1974, 1975, 1976b, 1977, 1978, 1984, 1985, 1987; Simon & Roscoe, 1981).

Simon contends that the performance of interest should be approached based on a program of research that is marked by what he calls "progressive iteration." By this he means to imply a program of research comprised of a series of dependent experiments rather than a collection of independent experiments. Within this series of experiments, later experiments build on those that precede them such that each new experiment is an extension of the experimental program as a whole. The series begins with a screening experiment that is designed primarily to order known and suspected determinants of the performance of interest by magnitude of effect. Typically, this first cut will use a fractional factorial study design in which all factors are represented by two levels. With such an experiment as the starting point, the research program proceeds to locate and isolate two-way, and possibly even three-way, interactions where the basic experiment suggests they may have appreciable magnitudes of effect. Some of the important factors may be nonlinear, in which case additional experiments are performed using three or more factor levels (i.e., central composite study designs) to describe their response surfaces. Thus, the screening experiment allows for a potentially large number of possible continuations. However, only one of these continuations will be realized, but it

should have been considered in advance and provisions made for it to be developed by extension from the original design.

In his approach, Simon emphasizes "economical multifactor design" (Simon, 1973). Accordingly, he argues that the minimum number of treatment conditions and data points be used to achieve the design objectives. The overall objective of his progressive iteration is to describe performance as a function of its determinants, thereby making the researcher's primary task one of parameter estimation rather than achieving statistical reliability. This last assertion, however, has resulted in much criticism of Simon's approach. Nevertheless, his approach is useful when one is more interested in understanding broad, less precise relationships across a large multivariate space than in obtaining highly reliable information about a few points in small segment of the experimental region (Simon, 1970). The former is clearly more relevant than the latter when considering capability tradeoff analyses. Additionally, Simon appears to have appreciated the importance of tradeoff analyses and provided for it in his research strategy (Simon, 1970):

> Engineers will find the regression model more useful than the ANOVA models more frequently used in human factors study. Regression equations can be used to…determine how equipment trade-offs should be made in order to optimize performance when one or more system parameters must be constrained (p. 5).

3. Isoperformance Curves

Simon's research strategy was aimed at solving the methodological problem of developing functional models of performance when there are many determinants of potential importance. In the 1980s, two other engineering psychologists, Robert Kennedy and Marshall Jones, working under sponsorship of the Army and Air Force, proposed using such formal models of performance to conduct the sort of tradeoff analyses that were conceptualized by DePuy and Bonder in their MPT capability tradeoff analysis (Jones, Kennedy, Turnage, Kuntz, & Jones, 1985; Jones, Kennedy, Kuntz, & Baltzley, 1987; Kennedy, Jones, & Baltzley, 1988, 1989; Kennedy, Jones, & Turnage, 1990; Kennedy & Jones, 1992; Jones, Turnage, & Kennedy, 1993; Jones & Kennedy, 1996; Jones, 2000; Jones, Kennedy, & Stanney, 2004). It appears that the work of Kennedy,

Jones and colleagues developed independently of that of DePuy and Bonder (e-mail from M. McCauley to R. Kennedy, January 29, 2010). Kennedy attributes the idea largely to prior research developing “isoemesis” curves (Figure IV-14) for ship movement induced motion sickness (McCauley & Kennedy, 1976) and a large scale program of studies (1979–1987) of simulator design for training purposes utilizing the Visual Technology Research Simulator (VTRS), formerly the Aviation Wide Angle Visual System (AWAVS), at the Naval Training Systems Center in Orlando, Florida (Lintern, Nelson, Sheppard, Westra, & Kennedy, 1981). The VTRS research was conducted primarily to identify simulator equipment features that best promoted acquisition of flying skills but also investigated variables associated with training and individual differences. A primary finding from a meta-analysis of the VTRS research was the important contribution made by individual differences to performance—subjects accounted for 50-80% of the variance in performance as compared to 10-25% for trials and 15-20% for equipment (Jones, Kennedy, Baltzley, & Westra, 1988; cited in Kennedy, Jones, & Baltzley, 1989). These results prompted Kennedy, Jones, and colleagues to contemplate the wisdom of trading off these main effects. Their basic logic was simple: in view of the fact that some pilots plainly have more ability than others, because repeated simulator sessions can be costly, and since expensive upgrades to equipment may have a minimal impact on performance, decision makers should focus on tradeoffs among these factors to obtain desired performance levels rather than trying to maximize any single factor (Turnage, Kennedy, & Jones, 1990).

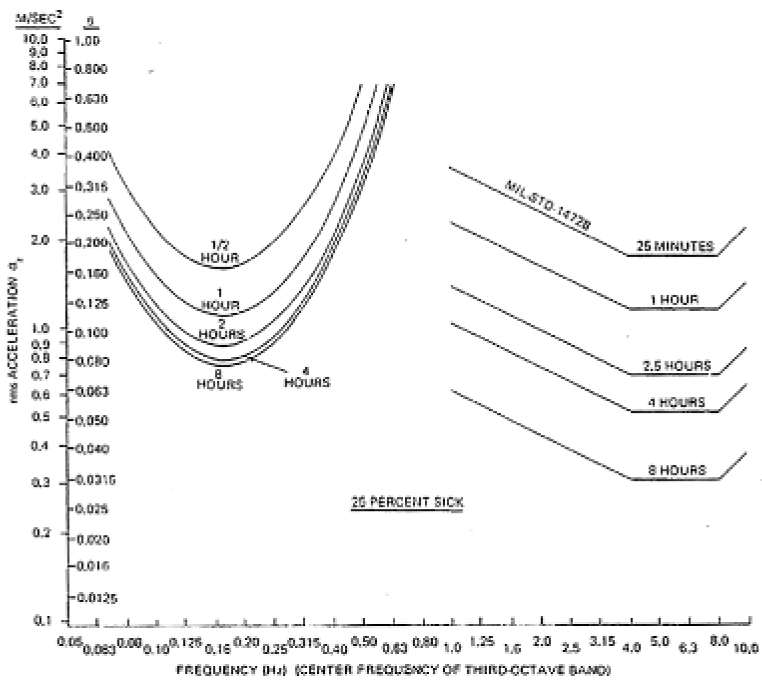


Figure IV-14. Equal motion sickness contours [From McCauley & Kennedy, 1976].

Given this genesis, the central reasoning behind their tradeoff methodology is the idea that, if one knows how performance varies as a function of its determinants, it is possible to derive equivalent combinations of the determinants. By equivalent, Kennedy and Jones mean that all such combinations produce the same specified level of performance—hence, the term "isoperformance" which they used to describe their approach to tradeoff analyses. Kennedy and Jones focused on personnel-training interactions and developed their technique primarily to generate tradeoff functions between personnel abilities and factors such as training time and training system effectiveness. However, they emphasize the generalizability of the isoperformance methodology to training, equipment, and manpower tradeoffs.

Isoperformance is ideally suited to problems that take the form of DePuy and Bonder's human/machine analysis, where a design engineer specifies a level of performance to be reached (e.g., the demand signal from the MPT front end analysis) regardless of the combinations of determinants used to do so (Jones, Kennedy, and Stanney, 2004). Isoperformance is inherently an applied methodology based on Simon's

(1996) notion of "satisficing," and so fixes the amount of performance at an acceptable level and trades off the determinants with respect to each other (de Weck & Jones, 2006). By implication, while the isoperformance methodology is itself atheoretic, it does presuppose a functional model between the determinants and the desired performance. The first step in the isoperformance technique is some data-analytic procedure based on a model, which may include common multi-variable statistical techniques like ANOVA or regression analyses. Such a model states the dependent variable(s) as a function of the determinants, parameters to be estimated, and error variations. Once the parameters are estimated, usually using least squares or maximum likelihood, the isoperformance technique requires that the user specify a *criterion* and a level of confidence, which is called the *assurance level*. The criterion is the level of performance desired by the user, and the assurance level is the probability of attaining that level of performance. Based on the user's choice of criterion and assurance level, the dependent variable is fixed and the resulting equation solved in terms of just the determinants; the determinants can now vary only in ways that will result in the same performance level. In the simple case of just two determinants, plots of every pair of values, each of which will produce the same level of performance, yields an isoperformance curve. Secondary criteria such as cost, safety, or feasibility are then used to identify a preferred solution(s) on the isoperformance curve (Jones & Kennedy, 1992; Jones & Kennedy, 1996; Jones 2000). In the context of DePuy and Bonder's human/machine analysis, such secondary criteria would include system proactive MPT constraints derived from the macro MPT planning analysis.

4. Illustrative Example: The Main Tank Gunner

DePuy and Bonder provide a conceptual process to interface MPT supply and demand, both at the macro, multi-system level and the micro, system-specific level. At the level of an individual system, the MPT supply and demand interface is accomplished through the human/machine analysis, which includes the MPT capability tradeoff analysis as a critical component. However, DePuy and Bonder leave the details of the MPT capability analysis to be worked out by future research efforts. While there is no indication that DePuy and Bonder were aware of the work of either Simon or Kennedy

and Jones, it is readily evident that the latter's isoperformance approach will accomplish the purpose of the MPT capability tradeoff analysis. To help illustrate this, what follows is a hypothetical demonstration of Kennedy and Jones' isoperformance approach using the example of the main tank gunner discussed by DePuy and Bonder.

You will recall that the MPT front end analysis hierarchically decomposes the functions allocated to the human in the human/machine tradeoff analysis into the supporting tasks, subtasks, and skill levels required to achieve the desired system performance. Thus, for example, the MPT front end analysis might produce the hierarchical structure shown in Figure IV-15 for some combat vehicle system. For the job of main tank gunner, we could have the following hierarchical decomposition:

- *Function*: Fire the gun at a moving target
- *Task*: Aim the gun
- *Subtask*: Track the target
- *Skill level*: Track with a 0.60 mil error or less

The skill level requirement is derived from the system performance requirement that the single shot hit probability be 0.75 or higher. The origin of the skill requirement highlights the point that the MPT front end analysis is accomplished without explicitly considering the impact of quality of personnel or training issues on the system. It simply considers the performance level for each function that is necessary to obtain the desired overall system performance, irrespective of how those functional performance levels might be achieved (DePuy & Bonder, 1982). To answer the question of "how," we turn to the MPT capability tradeoff analysis and Kennedy and Jones isoperformance approach.

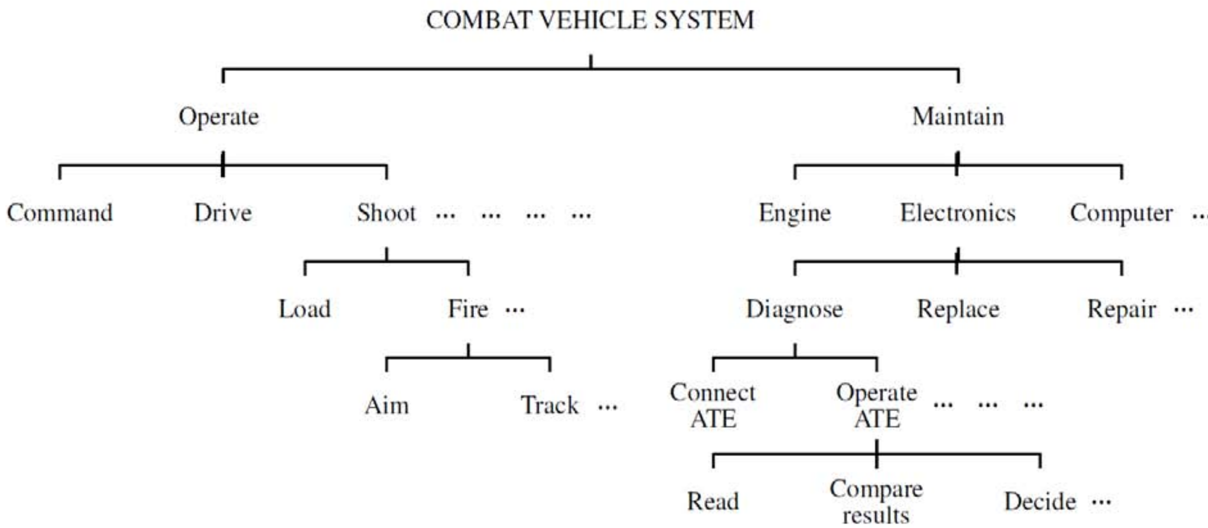


Figure IV-15. Function, task hierarchy for a hypothetical combat vehicle system [From DePuy & Bonder, 1982].

An isoperformance analysis is based on the proposition that one knows how performance varies as a function of its determinants. Since this is a hypothetical tradeoff analysis, we will use an empirical regularity, specifically the power law of practice, as a model to relate tracking error to personnel quality, expressed in terms of Armed Services Qualification Test (AFQT) score, and length of training. AFQT score is chosen as the measure of personnel quality because it has been shown to be both a measure of trainability and a predictor of performance in the Army (Horne, 1986). Accordingly, our model is of the following general form:

$$P = aT^{-b} \tag{2}$$

where P is our performance of interest (i.e., tracking error), a is an individual's initial level of performance, T is the amount of training in weeks, and b is a learning rate parameter. In practice, the parameters a and b are estimated based on empirical data. For our purposes here, we assume that initial performance level, a, is an arbitrary but monotonically decreasing function of AFQT score:

$$a = \frac{100}{\text{APFT score}} \tag{3}$$

where the range of AFQT scores is [21,100], corresponding to a basic qualification for enlistment of AFQT category IVA. Since law prohibits enlistments from AFPT category V, which encompasses AFPT scores in the range [0,9], we do not need to be concerned with the extreme case where *a* would be undefined because of division be zero. We also assume that the rate of learning parameter, *b*, is an arbitrary monotonically increasing function of APFT score:

$$b = \frac{\ln\left(\text{AFQT score}\right)}{10} \qquad (4)$$

The choice of functions is made solely to ensure appropriate scaling of the parameters, *a* and *b*, over the range of AFQT score. Consequently, personnel domain considerations (i.e., aptitude) can now be addressed in the power law through the values of the parameters *a* and *b*.

While DePuy and Bonder stipulate a tracking error requirement in their example of 0.60 mil, we will augment this requirement so that the format is more consistent with the manner in which the Defense Department specifies system requirements today. Accordingly, we assume that the requirement is for a maximum tracking error of 0.60 mil (threshold) / 0.40 mil (objective)—that is, tracking error must be less than 0.60 mil for the system to have military utility, but reducing the error to 0.40 mil will increase the military utility of the system. However, further reduction of tracking error below 0.40 provides no additional military utility—what is sometimes derisively referred to as "gold plating." We will also assume that there is a system proactive MPT training constraint, derived from the MPT planning analysis, for the length of training to be no greater than 10 weeks. Using the power law, the response surface between training time and individual aptitude is mapped out throughout the range of feasible values for *a*, *b*, and, *T*. The results are displayed in Figure IV-16, which is a surface chart showing tracking error as a function of AFQT score and training time. Color coding is used to show areas on the plot where performance exceeded the objective (green) or threshold (yellow) requirement. Areas on the plot failing to satisfy the threshold requirement are shown in red.

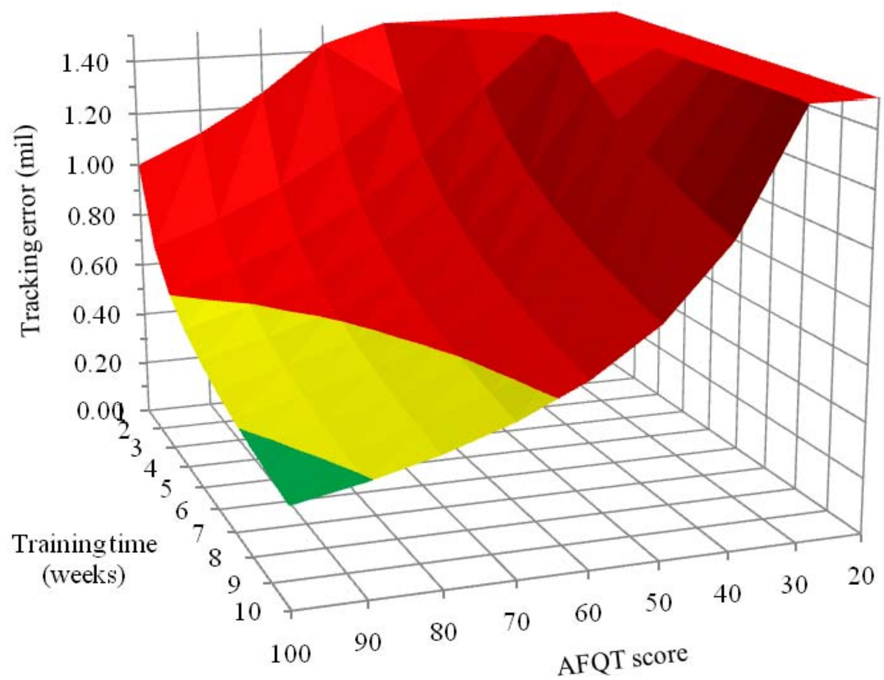


Figure IV-16. Surface plot of tracking error as a function of AFQT score (personnel domain of HSI) and training time (training domain of HSI).

To derive an isoperformance curve, we need to reformulate the power law so that training time is expressed as a function of AFQT score for some given fixed tracking error. The process to do this is relatively straightforward, but the steps are reviewed here for those who have not used algebra in a while. We begin with the following model formulation:

$$\hat{P} = aT^{-b} \tag{6}$$

where $\hat{P}$ is our fixed performance requirement and the right hand side is as defined for Equation 2. We next substitute in Equations 3 and 4 for the parameters *a* and *b* respectively:

$$\hat{P} = \left(\frac{100}{\text{AFQT score}}\right) T^{\frac{-\ln(\text{AFQT score})}{10}} \tag{7}$$

Rearranging terms gives us the following:

$$\hat{P}\left(\frac{\text{AFQT score}}{100}\right) = T^{\frac{-\ln(\text{AFQT score})}{10}} \tag{8}$$

Converting the exponential equation to its logarithmic equivalent:

$$\frac{-\ln\left(\text{AFQT score}\right)}{10}=\log_T\left[\hat{P}\left(\frac{\text{AFQT score}}{100}\right)\right] \tag{9}$$

We next change the logarithm from base T to base 10:

$$\frac{-\ln\left(\text{AFQT score}\right)}{10}=\frac{\log_{10}\left[\hat{P}\left(\frac{\text{AFQT score}}{100}\right)\right]}{\log_{10}\left(T\right)} \tag{10}$$

Rearranging terms again to isolate the term with T on the left hand side:

$$\log_{10}\left(T\right)=\frac{-10\cdot\log_{10}\left[\hat{P}\left(\frac{\text{AFQT score}}{100}\right)\right]}{\ln\left(\text{AFQT score}\right)} \tag{11}$$

Finally, we reformulate both sides of the equation as powers to the base 10, thereby leaving us training time as a function of AFQT score and required performance:

$$T=10^{\frac{-10\cdot\log_{10}\left[\hat{P}\left(\frac{\text{AFQT score}}{100}\right)\right]}{\ln\left(\text{AFQT score}\right)}} \tag{12}$$

Using Equation 12 to plot values of the two determinants, aptitude and training time, every pair of which produce the same level of performance, results in the isoperformance curves given in Figure IV-17. Since we do not have information on the distribution of our performance of interest, we can only consider the case where the assurance level is 50%, meaning we are ignoring the role of error in this analysis by limiting ourselves to considering combinations of determinants for which 50% of gunners will achieve the performance criterion. That concession is not of great consequence here since error is only part of the story and, for applied purposes, often not a very important part (Jones & Kennedy, 1996). Kennedy and Jones refer to a plot such as Figure IV-17 as an isoperformance readout. Nevertheless, Figure IV-17 looks quite similar to DePuy and Bonder’s conceptual capability tradeoff analysis shown earlier in Figure IV-12—so much so that it would be fair to say that the terms are synonymous.

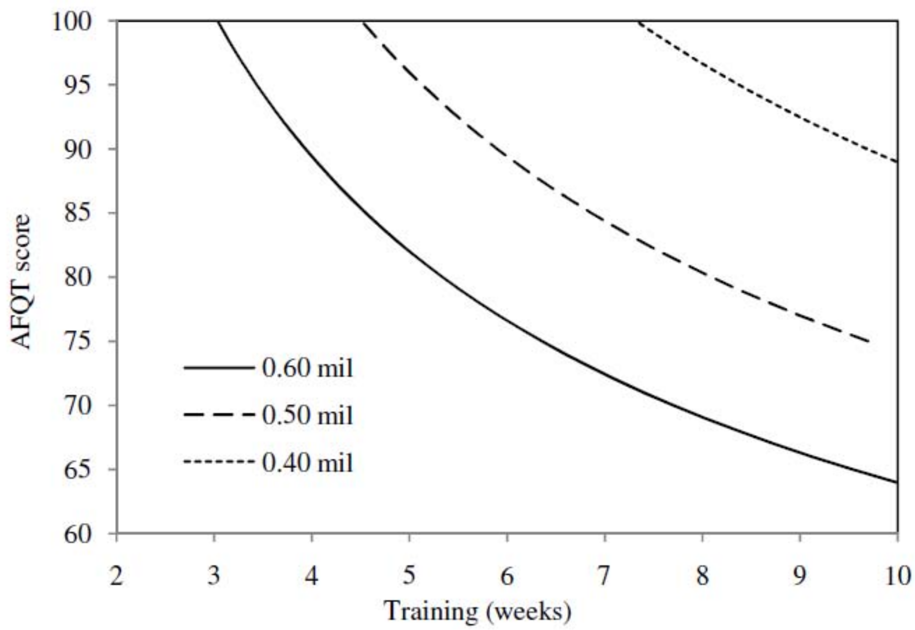


Figure IV-17. Isoperformance curves trading off aptitude (AFQT score) and training time with the criterion (tracking error) set at 0.60, 0.50, and 0.40 and level of assurance set at 0.50.

Isoperformance curves are read as tradeoff functions. For example, given a soldier with an AFQT score of 80, after 5.3 weeks of training the probability is 0.50 that they will track a target with a 0.60 mil error—that is, meet the threshold system performance requirement. A second soldier, who scores 65 on the AFQT, will take 9.5 weeks to reach the same performance level with a 0.50 probability. A drop of 15 points on the aptitude measure requires 4.2 weeks to make up. In other words, it takes two days of training to make up the reduction in performance indicated by a drop of one point in aptitude. Moreover, a soldier with an AFQT score of less than 64 will never be able to track a target with a 0.60 mil error with a probability of 0.50 given the constraint limiting training to less than ten weeks. The implication of this training constraint is that main tank gunners will necessarily need to be selected from AFQT Categories I and II. However, it is very likely that soldiers in these higher aptitude categories will be in demand for other systems. If it were determined that such higher quality soldiers will not be available in sufficient numbers, then it would become necessary to consider relaxing the training constraint, redesigning the task so that soldiers selected from lower AFQT

categories could achieve the performance criterion, or automating the task all together (i.e. revising the human/machine tradeoff analysis). The situation for the objective system performance requirement, tracking a target with a 0.40 mil error, is even more constrained, with only AFQT Category I soldiers capable of meeting the criterion with a probability of 0.50 given less than ten weeks of training. These latter types of considerations are issues for DePuy and Bonder's skill comparison analysis, but they are critically dependent on the information derived from the isoperformance curves.

E. ISOPERFORMANCE IN THE DESIGN OF COMPLEX SYSTEMS

Although the term "isoperformance" was originally coined in the area of human factors research, some have begun to advance the concept as an enabling methodology for a target-driven systems engineering process. The following sections are based in large part on a paper published by Olivier de Weck and Marshall Jones (2006) in the journal *Systems Engineering* applying the isoperformance approach to the design of complex systems. While their paper predominately focuses on the design of a satellite system, the methodology should generalize to other systems in which one or more humans contribute to overall system performance.

1. Isoperformance as a System Design Philosophy

Historically, system designers have sought the "best achievable" system performance, iteratively refining choices in the design space to maximize system objectives. The dominant mode of thought among such system designers is that of "forward analysis," where a vector of choices, $\mathbf{x}$, in the design space is uniquely mapped to an expected outcome in the objective space, $\mathbf{x} \mapsto \mathbf{J}(\mathbf{x})$. This paradigm has its roots in the mid 20th century when system performance was the prime driver for competitiveness and superiority. However, de Weck and Jones (2006) assert that a more natural mode of thought is, instead, to accept system performance that is "good enough" as determined based on contractually specified requirements, the need to achieve robust functionality at the lowest cost, etc. Such an approach results in desired performance levels becoming known quantities that can serve as targets for design engineers. This mode of thought is one of "inverse design," where a set of solutions are found in the design space that satisfy

as set of performance targets in the objective space, $\mathbf{J} \mapsto \mathbf{x}(\mathbf{J})$. The main challenge for the system designer lies in the nonuniqueness of the problem as there may be many design vectors, $\mathbf{x}_i$, that satisfy the vector of performance targets, $\mathbf{J}_{req}$.

De Weck and Jones (2006) offer isoperformance as a method for addressing this problem. They describe isoperformance as an inverse design method that first obtains a performance invariant set of system design solutions and subsequently reduces these to an efficient set when evaluated against other criteria such as cost and risk. Finding isoperformance solutions gives rise to three sequential issues. First, we must find ways to systematically search the design space for acceptable solutions. Second, once such solutions are obtained, we must find means to reduce the large set of alternatives to a smaller set that is suitable for presentation to decision makers. And third, criteria for selecting a particular solution in the reduced set must be discussed.

System performance requirements typically fall into one of three classes: "smaller-is-better" (SIB), "larger-is-better" (LIB), and "nominal is better" (Figure IV-18). The isoperformance approach assumes that desired performance targets, J_{req}, are known, meaning that the key performance objectives are captured as NIB and that they must be achieved first. The subset of solutions that strictly satisfy the NIB requirements are then extracted and analyzed for any potential engineering insight that might be gained from them. The next step is to evaluate the performance invariant solutions in terms of their SIB and LIB objectives and to filter the set of solutions that are non-dominated according to the SIB and LIB objectives. Finally, one or more designs are selected from this Pareto set for further consideration.

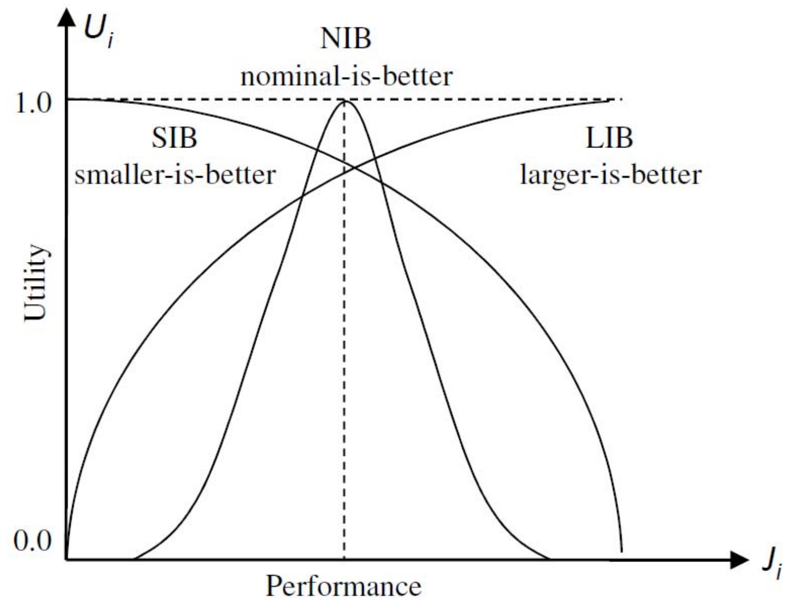


Figure IV-18. Normalized utility curves u_i for the i^{th} system objective J_i represented by monotonically decreasing (SIB), increasing (LIB) or concave function (NIB) [From de Weck & Jones, 2006].

2. Isoperformance Problem Formulation

a. Notation

To properly formulate the isoperformance problem, it is first necessary to establish some concepts and associated notation. A performance objective for a system is usually defined in terms of some statistical measure of system performance such as an average, maximum, minimum, root-mean-square, etc. Given $z(t)$ as the time-varying system performance output signal, the performance of the system in terms of the system objective can be written as some statistical measure or function of the output signal:

$$J_z = f(z) \tag{13}$$

where *z* depends on the settings of the design variables x_i, where $i = 1, \ldots, n$, as well as fixed parameters p_j:

$$z = h(x_i, p_j) \tag{14}$$

Recalling our main gunner example from Section IV-D4, the system performance output signal, $z(t)$, is tracking error, and the system objective, J_z, is described in terms of the root mean square error for tracking. Accordingly:

$$J_z = \|z\| = \mathrm{E}\left[z^T z\right]^{1/2} = \left(\frac{1}{T}\int_0^T z(t)^2\, dt\right)^{1/2} \tag{15}$$

where z is a function of training time, x_1, and AFQT score, x_2, and the fixed parameters in Equation 7.

An isoperformance requirement can then be formulated as

$$J_z\left(x_{iso,i}\right) \equiv J_{z,req} \qquad \forall\ i = 1,\ldots,n \tag{16}$$

whereby a two-sided tolerance band, τ, is allowed for practical and numerical reasons:

$$\tau: \ \left|\frac{J\left(x_{iso,i}\right) - J_{req}}{J_{req}}\right| \le \tau \tag{17}$$

Thus, $x_{iso,i}$ is any setting of the design variables such that the statistical measure of the output signal for system performance satisfies the specified system objective within some acceptable tolerance. In our example, $x_{iso,i}$ is any combination of training time and AFQT score that produce a root mean square tracking within some percent difference (τ) of the performance target (i.e., 0.60 mil error).

b. Step 1: Find the Performance Invariant Set

Find all $\mathbf{x}_{iso}^k \subseteq \mathbf{X}_{iso}$ such that for an objective vector $\mathbf{J}_z = \left[J_{z,1} \ldots J_{z,m}\right]^T$ the following constraints are satisfied:

$$\left|\frac{\mathbf{J}_z\left(\mathbf{x}_{iso}^k, \mathbf{p}\right) - \mathbf{J}_{z,req}}{\mathbf{J}_{z,req}}\right| \le \boldsymbol{\tau} \qquad \text{(Isoperformance constraints)} \tag{18}$$

$$\mathbf{g}(\mathbf{x}_{iso}^k, \mathbf{p}) \le 0,\ \mathbf{h}(\mathbf{x}_{iso}^k, \mathbf{p}) = 0 \qquad \text{(Feasibility constraints)} \tag{19}$$

$$x_{i,LB} \le x_{iso}^k \le x_{i,UB} \qquad \text{(Side bounds)} \tag{20}$$

In our example, we see that the isoperformance set for our threshold objective is described by the contour in Figure IV-17 labeled 0.60 mil. For the sake of simplicity, we approximate Equation 12 with a fitted power function so that the contour is described over the region of interest by the following equation:

$$\mathbf{X}_{iso}: \quad x_1 = 779684 \cdot \left(x_2\right)^{-2.71} \tag{21}$$

where x_1 is training time in weeks and x_2 is AFQT score. We extract from it a subset of three discrete isoperformance solutions (*A*, *B*, and *C*), *k* = 1, 2, 3:

$$\mathbf{x}_{iso}^{1} = \begin{bmatrix} 3.44 \\ 94.69 \end{bmatrix}, \quad \mathbf{x}_{iso}^{2} = \begin{bmatrix} 6.36 \\ 75.46 \end{bmatrix}, \quad \mathbf{x}_{iso}^{3} = \begin{bmatrix} 9.54 \\ 64.96 \end{bmatrix} \tag{22}$$

These satisfy the feasibility constraint that training time does not exceed 10 weeks:

$$x_1^k \le 10 \tag{23}$$

They also satisfy the following side bounds:

$$x_{1,LB}^k \ge 0 \tag{24}$$

$$x_{2,LB}^k = 21 \le x_2^k \le x_{2,UB}^k = 100 \tag{25}$$

The first side bound merely states that training time must be non-negative, while the second side bound restricts AFQT scores to those that are both technically achievable and consistent with public law.

c. Step 2: Find the Efficient Subset

Find all $\mathbf{x}_{iso}^{l*} \in \mathbf{X}_{iso}^{*} \subseteq \mathbf{X}_{iso} \subseteq \mathbb{R}^n$ such that for a vector of secondary (cost and risk) objectives $\mathbf{J}_{cr} = \left[J_{cr,1} \ldots J_{cr,r} \right]^T$, there exists no other feasible design vector $\mathbf{x} \in \mathbf{X}_{iso} \subseteq \mathbb{R}^n$ such that

$$\mathbf{J}_{cr}\left(\mathbf{x}\right) \le \mathbf{J}_{cr}\left(\mathbf{x}_{iso}^{*}\right) \qquad \text{(Non-inferiority constraint)} \tag{26}$$

$$J_{cr,j}\left(\mathbf{x}\right) \le J_{cr,j}\left(\mathbf{x}_{iso}^{*}\right) \qquad \text{(Component-wise non-inferiority constraint)} \tag{27}$$

for all $j = 1, \ldots, r$ with strict inequality holding for at least one *j*. Any feasibility constraints or side bounds on $\mathbf{x}_{iso}^{*}$ have already been satisfied by virtue of step 1.

In our main gunner example, the risk of a design might relate to the aptitude sensitivity of the tradeoff decision. Accordingly, we introduce the following risk objective:

$$J_{cr,1} = \left| f'\left(x_2\right) \right| \text{ where } f' = \frac{\partial x_1}{\partial x_2}. \tag{28}$$

The cost, on the other hand, might be captured by aptitude since increased incentives are required to attract higher quality recruits: $J_{cr,2} = \log\left[\beta x_2\right]$. By setting $\beta = 0.05$ we normalize the cost of $x_{2,LB}$ to unity. We can evaluate the three isoperformance designs, *A*, *B*, and *C* in terms of $\mathbf{J}_{cr}$ and obtain

$$J_{cr}\left(\mathbf{x}_{iso}^{1}\right) = \begin{bmatrix} 0.10 \\ 0.65 \end{bmatrix}, \quad J_{cr}\left(\mathbf{x}_{iso}^{2}\right) = \begin{bmatrix} 0.23 \\ 0.56 \end{bmatrix}, \quad J_{cr}\left(\mathbf{x}_{iso}^{3}\right) = \begin{bmatrix} 0.40 \\ 0.49 \end{bmatrix} \tag{29}$$

All three solutions are nondominated or efficient points according to our non-inferiority constraint. The first solution is the low risk, high cost option, the second is a compromise between risk and cost, and the third is the high risk, low cost option.

d. Step 3: Select the Final Design

Select $\mathbf{x}_{iso}^{**} \in \mathbf{X}_{iso}^{*} \subseteq \mathbf{X}_{iso} \subseteq \mathbb{R}^n$ according to non-quantified objectives or criteria and stakeholder considerations.

It should be noted that steps 1 and 2 could be collapsed into a single multi-objective optimization problem with two-sided inequality constraints on performance. For simple design problems, the results of either the isoperformance method or the all-in-one optimization should be equivalent. However, for complex design problems, de Weck and Jones claim two advantages of the isoperformance method. First, engineering insight is obtained by solving the problem in steps, and inspecting and visualizing the intermediate results. And second, as the tolerance is decreased, $\tau \to 0$, the isoperformance constraint can become difficult to solve using standard multiobjective techniques.

3. Deterministic Versus Empirical Isoperformance Models

It is self-evident that isoperformance requires some mathematical model that relates the vector of design variables, $\mathbf{x}$, to both performance and cost/risk objectives, $\mathbf{J}_z$ and $\mathbf{J}_{cr}$, respectively. The model can be derived from first principles and known empirical regularities (as was done in Equation 7) or obtained from experimental or field observation. In the case of the former, which is the deterministic case, the model is developed and implemented directly. In the latter case, an empirical model with embedded uncertainty must be developed first. This, in turn, requires a dataset that relates experimental factors $(\mathbf{x})$ to observed outcomes $(\mathbf{J})$. While it is often possible to make use of deterministic, physics-based models when considering the performance of technological systems, empirical models will likely predominate when $\mathbf{J}_z$ incorporates the performance of humans. Figure IV-19 illustrates these two general cases, and it is worth noting the isoperformance algorithms used to extract the performance invariant set, $\mathbf{X}_{iso}$, are the same in both cases.

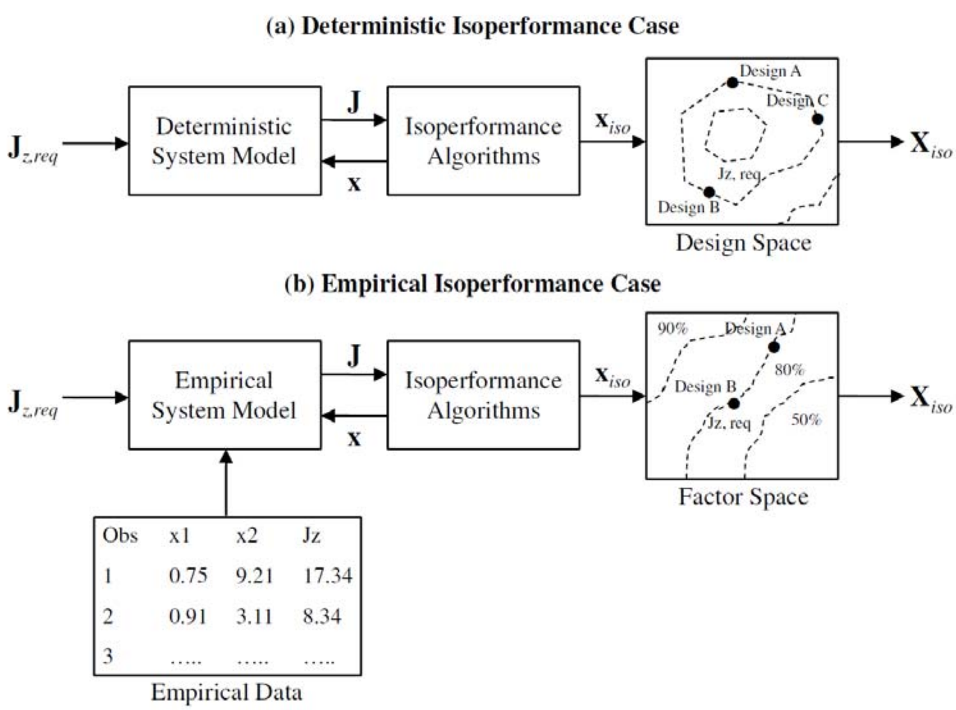


Figure IV-19. Deterministic versus empirical modeling approaches [From de Weck & Jones, 2006].

F. COUPLING ISOPERFORMANCE WITH UTILITY ANALYSIS

1. Problem Statement

As it is usually presented, the isoperformance approach assumes that desired performance targets are known. Nevertheless, an inevitable question is whether we can relax that assumption. Kennedy and Jones published a short proposal in a *SAE Technical Paper* (Kennedy, Jones, & Turnage, 1990) describing a strategy to couple their isoperformance methodology with utility analysis. Given their work from 1979 to 1987 studying simulator design for training purposes, Kennedy and Jones' paper primarily focuses on the application of the isoperformance method to a large simulator sickness database to create isoperformance curves for controlling the adverse consequence of exposure to simulators. After discussing the importance of the isoperformance methodology in terms of tradeoff analysis, Kennedy and Jones take the next logical step of considering the role of isoperformance in decision analysis:

> …system development requires making human resource decisions that have substantial bottom line implications…decision makers have had great difficulty in determining the effectiveness or value of human resource decisions in the same manner that production and engineering decisions can be evaluated. […] One overlooked tool that has direct application to this problem is *utility analysis* [emphasis added] (p. 3).

While isoperformance provides information on equivalent combinations of determinants that produce the same level of performance, Kennedy and Jones posit that utility analysis could guide the selection of the target performance level or criterion. In their thinking, this last step, coupling the choice of target performance level to some utility (or value) trade space, is necessary if human performance considerations are to be explicitly included within the larger context of total system analyses and tradeoffs.

Kennedy and Jones hint towards the solution of this problem in their discussion of a "meta-model" using isoperformance and utility analysis, but they do not explicitly describe what general form such a meta-model should take. To begin to address this issue, we need to make some assumptions about the nature of a problem that might involve coupling isoperformance and utility analysis. Such a problem will likely involve a system, the latter being broadly understood to mean a complex set of human and/or

technological components that interact to achieve a desired function. Performance is a quantitative measure of how well this function is executed, and utility is attained based on the quality of this performance. The quality of performance is also a function of both individual aptitude and training. Moreover, some selection criteria exist for identifying individual aptitude and utility benefits are gained by establishing selection criteria thresholds independent of the resulting performance. Similarly, utility benefits can be accrued for training interventions independent of the resulting impacts on the quality of performance. It is necessary that some utility be attained from selection policies and training independent of performance quality; otherwise, there is no need to consider either selection or training explicitly in a utility analysis. Instead, one would simply consider the utility of the resulting performance. However, it is often the case that there are significant organizational and logistical considerations involved in setting selection criteria and training times. For example, the length of training may directly impact personnel availability as well as have indirect implications for the availability of training resources for other systems. When such considerations exist, it is preferred to approach the problem from the perspective of multiple criteria optimization so that all dimensions of the solution space are adequately considered and the nature of tradeoffs between dimensions is understood.

Attempts to mathematically model multiple criteria optimization problems generally assume a similar approach: decision dimensions are mapped onto a unitless scale of goodness and combined together into an overall measure of goodness. Many of these methods are Archimedian or weight-based approaches that are very difficult to implement for realistic problems that comprise many weights. Such methods typically require many iterations of the choice of weights and often provide inadequate guidance on how to converge on the right set of weights. Additionally, in the case when objectives are nonlinearly dependent on the decision variables, the iterative development of weights can be quite problematic. Even the choice of utility functions in utility theory can be interpreted as involving some implicit choice of weights. Moreover, while it might be possible to determine the correct weights in a certain neighborhood of the solution space, there is no guarantee those weights will remain valid in a different neighborhood. While

preemptive approaches avoid the choice of weights by explicitly ordering competing objectives, they suffer in that they allow higher priority objectives to dominate lower priority objectives and often explore only a portion of the potential solution space (Messac, 1996; Messac, Gupta, & Akbulut, 1996).

A framework for a meta-model combining isoperformance and utility analysis needs to allow the problem formulation to retain its multiobjective character, by which we mean that a decision maker should not be forced to form a weighted sum of several criteria. It should also allow for the possibility of deliberate imprecision in the statement of preferences. Messac and colleagues (1996) suggest that goal programming possesses the first feature, while *physical programming* possesses both. A defining characteristic of physical optimization is that it exploits the availability of important information regarding the physical meaning of many of the problem parameters and criteria so as to remove the decision maker from the highly subjective task of choosing weights. Consequently, we will consider physical programming as a framework for specifying our meta-model coupling isoperformance and utility analysis.

2. Physical Program Problem Formulation

a. Notation and Model Specifications

We denote the decision variable vector as $\mathbf{x}$ and the i^{th} generic decision criterion as $g_i(x)$. The value of the criterion under consideration, g_i, is on the horizontal axis, and the function that will be minimized for that criterion, z_i, called the class function, is on the vertical axis. By convention, a lower value of the class function is better than a higher value, and the ideal value of the class function is zero.

In physical programming, preference generically falls under three classes, each comprising two cases. As depicted in Figure IV-20, the preference classes are referred to as follows: 1) class-1: smaller-is-better (SIB), 2) class-2: larger-is-better (LIB), and 3) class-3: nominal-is-better (NIB). The two cases, hard and soft, refer to the sharpness of the preference. The soft class functions provide a means for decision makers to express ranges of differing levels of preference for each criterion, resulting in a

deliberately imprecise framework. In contrast, hard class functions are used only when one cares to stay within some limits. The aggregate objective function (to be minimized) includes only soft class functions; hard class functions simply act as constraints.

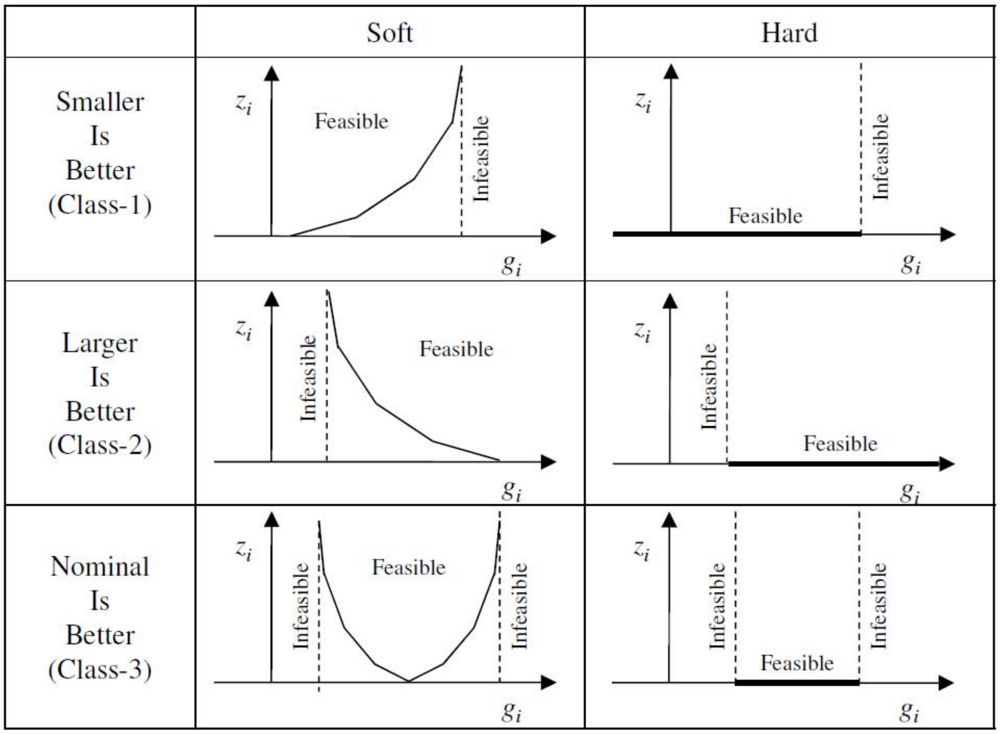


Figure IV-20. Preference function classification in physical programming [After Messac, 1996].

For the soft case, the physical programming lexicon comprises terms that characterize the degree of desirability of up to 10 ranges. To illustrate the physical programming lexicon, consider the soft case of class-1 (class-1S) shown in Figure IV-21. The ranges are defined as follows, in order of preference:

- *Ideal* $\left(g_i \le t_{i1}^{+};\ \text{range-1}\right)$: A range over which every value of the criterion is ideal (the most desirable possible). Any two points of that range are of equal value to the decision maker.
- *Desirable* $\left(t_{i1}^{+} \le g_i \le t_{i2}^{+};\ \text{range-2}\right)$: An acceptable range that is desirable.
- *Tolerable* $\left(t_{i2}^{+} \le g_i \le t_{i3}^{+};\ \text{range-3}\right)$: An acceptable, tolerable range.

- *Undesirable* $\left(t_{i3}^{+} \le g_i \le t_{i4}^{+};\ \text{range-4}\right)$: An acceptable range that is undesirable.
- *Highly undesirable* $\left(t_{i4}^{+} \le g_i \le t_{i5}^{+};\ \text{range-5}\right)$: An acceptable range that is highly undesirable.
- *Unacceptable* $\left(g_i \ge t_{i5}^{+};\ \text{range-6}\right)$: A range of values that the generic criterion may not take.

The parameters/targets t_{i1}^{+} through t_{i5}^{+} are *physically meaningful* values that are specified by the decision maker to quantify the preference associated with the i^{th} criterion. They may be derived from organizational policies, planning forecasts, or system specifications. These parameters delineate the desirability ranges within each criterion, thereby determining the shape of the class function.

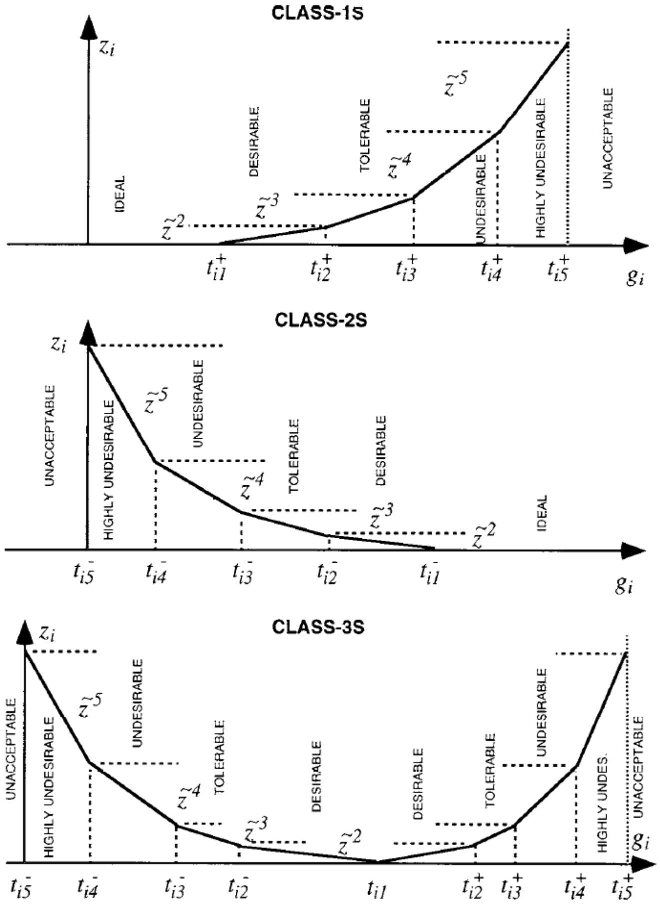


Figure IV-21. Class function regions for the i^{th} generic criterion [After Messac, Gupta, & Akbulut, 1996].

It is worth pointing out that the ideal range is defined such that any two points within this range are of equal value. Thus, the class function is explicitly indifferent to values in the ideal range. This behavior is complementary to the manner in which the Defense Department currently specifies system requirements. In the Defense Department lexicon, performance must exceed a certain minimum threshold requirement, t_{i5}^{+}, for the system to have military utility. Failure to achieve this threshold calls into question whether the acquisition program should be continued or cancelled. Improvements in performance beyond the threshold are desirable up to some maximum value or objective, t_{i1}^{+}, after which further improvements in performance provide no additional military utility (i.e., performance beyond t_{i1}^{+} would qualify as "gold plating" in the Defense Department vernacular). Consequently, setting physically meaningful targets should be relatively straightforward for systems engineers and program managers working on Defense Department programs.

For a given criterion, once the decision maker decides to which class the criterion belongs and chooses the range targets, the intracriterion preference statement is complete. Since this a multi-objective problem, there must also be intercriteria preferences. Physical programming makes use of the "One versus Others" criteria rule (OVO rule) as an implicit form of intercriteria preference. To understand the OVO rule, consider the following two options concerning class function values:

- Option 1: Full reduction for one criterion across a given range (range k, k = 3, 4, 5).
- Option 2: Full reduction of all of the other criteria across the next better range [region (k – 1)].

The OVO rule states that Option 1 shall be preferred over Option 2. In other words, it is considered better for one criterion to travel across, for example, the tolerable region than it is for all the other criteria to travel across the desirable region.

The next step is to develop the soft class functions. All functions must have the following properties:

1) A lower value of the class function is preferred over a higher value thereof.

2) A class function is strictly positive.
3) A class function is continuous, piecewise linear, and convex.
4) The value of a class function, z_i, at a given range-intersection is the same for any class-type.
5) The magnitude of a class function's vertical excursion across any range must satisfy the OVO rule.

The justifications for the preceding properties are provided in Messac, Gupta, and Akbulut (1996). Based on these properties, they develop the following relationships for the i^{th} criterion and the s^{th} range intersection:

1) From property (4):

$$z^s \equiv z_i\left(t_{is}^+\right) \equiv z_i\left(t_{is}^-\right) \quad \forall i;\ \left(2 \le s \le 5\right);\ z^1 \equiv 0 \tag{30}$$

2) The change in z_i that takes place as one travels across the s^{th} range, $\tilde{z}^s$, is given by:

$$\tilde{z}^s \equiv z^s - z^{s-1} \quad \left(2 \le s \le 5\right);\ z^1 \equiv 0 \tag{31}$$

3) To enforce the OVO rule:

$$\tilde{z}^s = \beta\left(n_{sc} - 1\right)\tilde{z}^{s-1} \quad \left(3 \le s \le 5\right);\ n_{sc} > 1;\ \beta > 1 \tag{32}$$

where n_{sc} is the number of soft criteria and β is a convexity parameter. To apply Equation 32, $\tilde{z}^2$ is set to a small positive value such as 0.01.

4) The length of the s^{th} range on the i^{th} criterion is defined as:

$$\begin{aligned} \tilde{t}_{is}^+ &= t_{is}^+ - t_{i(s-1)}^+ \quad \left(2 \le s \le 5\right) \\ \tilde{t}_{is}^- &= t_{is}^- - t_{i(s-1)}^- \quad \left(2 \le s \le 5\right) \end{aligned} \tag{33}$$

With this quantity, the magnitude of the slopes of the generic i^{th} criterion takes the form:

$$\begin{aligned} w_{is}^+ &= \frac{\tilde{z}^s}{\tilde{t}_{is}^+} \quad \left(2 \le s \le 5\right) \\ w_{is}^- &= \frac{\tilde{z}^s}{\tilde{t}_{is}^-} \quad \left(2 \le s \le 5\right) \end{aligned} \tag{34}$$

As suggested by Equation 34, the slopes vary across ranges and criteria. Once the slopes are known, the convexity requirement in verified by the relationship:

$$\tilde{w}_{\min} = \min_{i,s}\left\{\tilde{w}_{is}^{+}, \tilde{w}_{is}^{-}\right\} > 0 \quad \begin{cases} (2 \le s \le 5) \\ i: \text{ soft criteria} \end{cases} \tag{35}$$

where

$$\left.\begin{aligned} \tilde{w}_{is}^{+} &= w_{is}^{+} - w_{i(s-1)}^{+} \\ \tilde{w}_{is}^{-} &= w_{is}^{-} - w_{i(s-1)}^{-} \\ w_{i1}^{+} &= w_{i1}^{-} \end{aligned}\right. \begin{cases} (2 \le s \le 5) \\ i: \text{ soft criteria} \end{cases} \tag{36}$$

It should be noted that the quantities $\tilde{w}_{is}^{+}$ and $\tilde{w}_{is}^{-}$ calculated in Equation 36 are exactly the weights that will be used in the physical optimization model of the class functions. Equation 35 states that so long as the weights are positive, the class function will be piecewise linear and convex. Moreover, convexity can always be satisfied by simply increasing the magnitude of the convexity parameter, β, in Equation 32.

b. Physical Programming Weight Algorithm

The following algorithm evaluates the weights that are to be used in the physical program model of the class functions:

1) Initialize:

   $\beta = 1.1;\ w_{i1}^{+} = 0,\ w_{i1}^{-} = 0;\ \tilde{z}^{2} = 0.1;\ i = 0;\ s = 1;\ n_{sc} =$ number of soft criteria

2) Set $i = i + 1$

3) Set $s = s + 1$

   Evaluate in sequence $\tilde{z}^{s},\ \tilde{t}_{is}^{+},\ \tilde{t}_{is}^{-},\ w_{is}^{+},\ w_{is}^{-},\ \tilde{w}_{is}^{+},\ \tilde{w}_{is}^{+},\ \tilde{w}_{\min}$

   If $\tilde{w}_{\min} < 0.01$, then increase β by 0.01 and go to step 2.

4) If $s \neq 5$ go to step 3.

5) If $i \neq n_{sc}$, go to step 2.

Using the weights obtained from the above algorithm, which were derived without any decision maker input other than the range targets $\left(\text{i.e., } t_{is}^{+},\ t_{is}^{-}\right)$, it is now possible to formulate the physical programming problem model.

c. Physical Programming Problem Model

Building on the previous development, the physical programming problem statement takes the following form:

$$\min_{d_{is}^-,\, d_{is}^+,\, x} \quad J = \sum_{i=1}^{n_{sc}} \sum_{s=2}^{5} \left(\tilde{w}_{is}^- d_{is}^- + \tilde{w}_{is}^+ d_{is}^+ \right) \tag{37}$$

subject to

$$\left.\begin{aligned} & g_i - d_{is}^+ \le t_{i(s-1)}^+ \\ & g_i \le t_{i5}^+ \\ & d_{is}^+ \ge 0 \end{aligned}\right\} \quad \left(\text{for all } i \text{ in classes 1S and 3S, } i = 1,\ 2, \ldots, n_{sc},\ s = 2, \ldots, 5\right)$$

$$\left.\begin{aligned} & g_i + d_{is}^- \ge t_{i(s-1)}^- \\ & g_i \ge t_{i5}^- \\ & d_{is}^- \ge 0 \end{aligned}\right\} \quad \left(\text{for all } i \text{ in classes 2S and 3S, } i = 1,\ 2, \ldots, n_{sc},\ s = 2, \ldots, 5\right)$$

and

$$g_j \le t_{j,\max} \qquad \left(\text{for all } j \text{ in class 1H, } j = 1,\ 2, \ldots, n_{hc}\right)$$

$$g_j \ge t_{j,\min} \qquad \left(\text{for all } j \text{ in class 2H, } j = 1,\ 2, \ldots, n_{hc}\right)$$

$$t_{j,\min} \le g_j \le t_{j,\max} \qquad \left(\text{for all } j \text{ in class 2H, } j = 1,\ 2, \ldots, n_{hc}\right)$$

$$\mathbf{x}_{\min} \le \mathbf{x} \le \mathbf{x}_{\max}$$

where d_{is}^- and d_{is}^+ are deviational variables, $\mathbf{x}$ is the decision variable vector, $g_i = g(\mathbf{x})$, and n_{hc} denotes the number of hard criteria. Given the assumptions in our problem statement in Section IV-F1, one of the three soft class functions will apply to individual aptitude (i.e., personnel), training, and performance criteria derived from the isoperformance analysis. Thus, included among the n_{sc} soft criteria in Equation 32 are the following:

$$\begin{aligned} g_1 &= g_{training}(\mathbf{x}) \\ g_2 &= g_{personnel}(\mathbf{x}) \\ g_3 &= g_{iso}(g_1, g_2) \end{aligned} \tag{38}$$

Accordingly, the isoperformance analysis has been coupled with utility analysis in a physical optimization meta-model.

3. Example

A simple problem is chosen in an effort to introduce the methodology without unduly obscuring its key features. Recalling our main gunner example from Section IV-D4, the quantities of interest—or criteria—for the person performing the human/machine analysis are as follows:

Training:

$$g_1 \equiv x_1 \tag{39}$$

Aptitude:

$$g_2 \equiv x_2 \tag{40}$$

Tracking error:

$$g_3 \equiv p_{iso} = \frac{100}{x_2} \cdot x_1^{\frac{-\ln(x_2)}{10}} \tag{41}$$

The decision variables are

$$\mathbf{x} = \{x_1, x_2\} \tag{42}$$

where x_1 is the length of training in weeks and x_2 is AFQT score. Various wishes are expressed by the pertinent stakeholders as follows:

- Training: Shorter is better to decrease resource utilization and ownership costs.
- Aptitude: The lower the score, the better, as the pool of potentially available soldiers in increased.
- Tracking error: Lower is better between the threshold and objective values as specified in the system requirement.

Consequently, 1-S class functions are chosen for all three criteria. The range limits that delineate degrees of desirability as expressed by the stakeholders are reported in Table IV-2.

Table IV-2. Physical programming region limits table.

i^{th} criteria	Class type	t_{i1}^{+}	t_{i2}^{+}	t_{i3}^{+}	t_{i4}^{+}	t_{i5}^{+}
x_1	1-S	2	5	8	9	10
x_2	1-S	21	30	49	64	92
p_{iso}	1-S	0.40	0.45	0.50	0.55	0.60

The weights generated using the physical programming weight algorithm and the optimal decision results are presented in Table IV-3. We see that the most utility is attained by accepting borderline tolerable/undesirable values for training time and tracking error and a highly undesirable value for aptitude (corresponding to AFQT category I and II personnel). It is worth noting that Messac and colleagues' lexicon for describing the target ranges sounds a bit odd as applied in this problem context, particularly for the tracking error criterion. We maintained it here for the sake of consistency, but we suggest that other Likert-type ordinal listings of preference might be chosen based on the problem context. That issue notwithstanding, this example clearly demonstrates an approach for selecting isoperformance thresholds based on the objective of optimizing overall utility.

Table IV-3. Physical programming optimization results.

i^{th} criteria	Weights				Value
	$\tilde{w}_{i2}^{+}$	$\tilde{w}_{i3}^{+}$	$\tilde{w}_{i4}^{+}$	$\tilde{w}_{i5}^{+}$	
x_1	0.03	0.10	1.48	4.88	8
x_2	0.01	0.01	0.09	0.12	80
p_{iso}	2.00	6.04	24.28	97.61	0.50

G. BRINGING IT ALL TOGETHER

To illustrate the key points in this chapter, we will use the following notional systems decision problem: the Transportation Security Administration (TSA) wants to acquire new technology to enhance the effectiveness of airline passenger screening. The core of this example is an HSI trade space activity developed for use in one of the Naval Postgraduate School's HSI certificate courses. My intent in using this example is to help novice HSI practitioners visualize the process through which HSI tradeoff considerations can be addressed during the early phases of systems engineering and management. Given the absence of a published case study that accomplishes this goal, we must content ourselves with a reasonable facsimile thereof. The example presented in the following section is tailored with the objective of trading off complexity of detail and realism for the benefit of a cleaner, holistic perspective of the scenario. This section borrows heavily from material presented in chapters 10–12 in *Decision Making in Systems Engineering and Management* (Parnell, Driscoll, & Henderson, 2008). Interested readers can refer to that text for more detail on specific topics.

1. Illustrative Example: Enhancing Airline Passenger Screening

There is no question that the airline security system failed in an extremely broad fashion on December 25, 2009. A 23-year-old Nigerian man boarded a Detroit bound flight with explosive material hidden in his underwear. He almost succeeded in killing 289 people but for the quick actions of the passengers. In the aftermath of this event, there have been calls for a variety of actions focused on improving intelligence collection, analysis, and dissemination as well as enhancing airline passenger screening. In the case of the latter, the TSA quickly opted for a materiel solution to the problem and announced a plan to deploy the latest in body scanning technologies. However, within days of their announcement, vigorous debate emerged over the use of "full body scanners" at airports and the ramifications for passenger privacy (Figure IV-22). Clearly, the initial problem statement of enhancing security was not a full statement of the problem from the perspective of all stakeholders.

Figure IV-22. Cartoon by Daryl Cagle from January 14, 2010, commenting on the heightened emphasis on security at the nation's airports following the unsuccessful attack by the underwear bomber [From Cagle, 2010].

2. Problem Definition

The following begins the work of developing our notional systems decision problem. An integrated product team working for the Department of Homeland Security was formed and tasked with fielding a new airline passenger screening system for use by the TSA. Given the pressure to rapidly field a solution, senior agency decision makers directed that the acquisition strategy focus on acquiring a non-developmental item (NDI) rather than pursuing the potentially lengthy development process required for a completely new solution. Accordingly, a request for proposals (RFP) was quickly published and two companies responded with their technical solutions: 1) Virtually Nude and 2) Magic Eyes.

Although the project manager wishes to move quickly to a source selection decision, their HSI consultant is equally fast to point out that acquiring a NDI-based

system represents a significant challenge because far more problems must be foreseen and forestalled rather than being detected and resolved during development. The HSI consultant suggests that establishing some requirements will help ensure that a NDI-based system is selected on a sound basis. They also emphasize that HSI requirements, in particular, will help avoid the risk that cheaper equipment will lead to higher human-related costs or poor overall system performance downstream. Given the persuasiveness of the HSI consultant's argument, the project manager acquiesces and the IPT goes about doing a quick stakeholder analysis and some requirements engineering, the results of which are summarized in Table IV-4. The constructive scores for the objective "minimize privacy impact" have the following values:

+1 Better than current system
0 Same as current system
- 1 Marginally worse than current system
- 2 Worse than current system
- 3 Significantly worse than current system

Table IV-4. Raw data matrix for airline passenger screening systems.

		Solutions		
Objectives	**Measures of effectiveness**	Virtually Nude	Magic Eyes	Baseline
Maximize reliability	Number of missed threats/1000 passengers screened	1*	3*	10
Maximize capacity	Number of passengers screened per hour	45	60	75
Minimize privacy impact	Constructive scale compared to current system†	-2	-1	0
Minimize manpower	Number of operators	3	2	2

*Reported by manufacturer with unspecified users.
†Constructive scores

The question for the HSI consultant to consider at this point is whether the data provided in Table IV-4 are adequate for proceeding to source selection. Since the scores reported for the reliability objective were obtained using unspecified users, the short answer to this question is that we do not know. Given that we generally consider a

system as being comprised of liveware (i.e., humans), hardware, and software, if the users involved in the testing resemble TSA employees, then the reported reliability scores are probably accurate. However, if the users involved in the testing differ from TSA employees in one or more attributes, then the TSA may get a very different system when they combine their employees with the hardware and software provided by either company. In this situation, the reported scores for reliability in Table IV-4 may not be very accurate, in which case there is a need to do some type of experimentation, whether using modeling and simulation or human-in-the-loop testing, to generate more realistic estimates of system performance. This, in turn, will inevitably lead to the issue of design of experiments, a topic we will discuss in more detail momentarily.

3. Defining the HSI Solution Space

As part of their analysis, the HSI consultant pays a visit to the TSA human resources director. There they learn that the TSA administers an aptitude assessment, *Screener-IQ*, to potential employees. A minimum score of 60 is required for employment. Aptitude scores of hires are normally distributed with a mean score of 100 and standard deviation of 10. All employees that are hired complete a minimum of six hours of generic classroom instruction on the existing screening system before beginning practical training until they are proficient with the system. It currently takes approximately 40 total hours before new hires reach proficiency. Because the TSA experiences a high rate of turnover, there is also a desire to minimize training time for any replacement screening system. Moreover, the TSA cannot accommodate any training program that is greater than 54 hours given the currently available training budget and resources.

Based on their visit with the human resources director, the HSI consultant can now augment the earlier top-level requirements engineering by providing supporting requirements for the personnel and training domains of HSI. Such requirements might take the following general form:

- *Personnel domain*: The system shall be operable by users with a minimum *Screener-IQ* aptitude score of 80 (threshold) / 60 (objective).

- *Training domain*: Initial training to use the system shall be no greater than 40 hours (threshold) / 20 hours (objective).

These requirements are themselves laden with stakeholders' (e.g., human resources director) values and could conceivably be described in terms of degrees of preference—a point that will be taken up later. However, for the moment, the HSI requirements are more easily thought of as providing MPT constraints on the potential solution space. In the case of the personnel domain, it was decided, at a minimum, to accommodate users who are at least two standard deviations below the mean, but it is preferred to accommodate all users who could potentially be hired (i.e., achieve the minimum *Screener-IQ* score of 60). Likewise, for the training domain, screener training time cannot exceed the available training hours, but it is preferred that training time be no longer than that of the legacy system. Moreover, decreasing the training time relative to the legacy system would provide significant long-term logistical savings. These types of MPT issues and constraints also feed into any subsequent design of experiments, providing important information regarding the necessary scope of experiments. Such scope considerations should include the range of individual aptitudes, training times, and training media that should be considered.

4. Evaluating Solutions

Since the two responses to the RFP both report reliability for their respective system without providing any description of the corresponding users, the HSI consultant recognizes the need to quickly and economically develop some data to help inform the systems decision process. The HSI consultant turns to design of experiments (DOE) as a means to simultaneously study the individual and interactive effects of several factors, thereby keeping the number of experimental replications to a minimum. Design of experiments addresses the basic question, "What is the average outcome (effect) when a factor is moved from a low level to a high level?" (West, 2008, p. 333). Since more than one factor is often of some importance in a complex system, efficiencies are gained by moving combinations of factors simultaneously.

Let us take a moment to consider an experimental design for evaluating the new airline passenger screening systems described above. The HSI consultant has initially identified three key factors that might influence performance: individual aptitude, length of training, and equipment. These factors correspond to the personnel, training, and human factors engineering domains of HSI respectively. The HSI consultant has also determined approximate levels of interest for each factor: aptitude scores ranging from 60 to 100, training times ranging from 10 to 50 hours, and the two equipment designs being considered. The primary measure of effectiveness is the "rate of missed threats." This design can be displayed as a cube with the three factors, aptitude score (*x*), training time (*y*), and scanner (*z*), shown with both low (–1) and high (+1) levels corresponding to the lower and upper levels of interest (see Figure IV-23). Eight conditions, or design points (DP), are possible in this design: DP_1 where *x* and *y* are set to their lower levels and *z* is set to its higher level, DP_2 where *x* and *z* are set high and y is set low, and so on as shown in Figure IV-23. This type of experimental design is called a 2^3 factorial design since it is based on three factors and each factor has two levels. This also tells the experimenter how many design points exist: $2^3 = 2 \times 2 \times 2 = 8$.

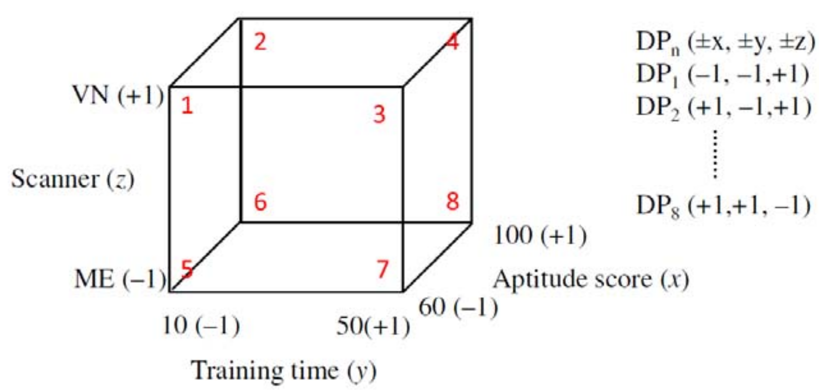


Figure IV-23. Three-factor, two-level design.

An additional strength of design of experiments is that it allows the experimenter to understand the interaction effects between factors. These interaction effects show the synergistic relationships between factors by measuring how the effect of one factor may

depend on the level of one or more other factors. A full 2^3 factorial design allows for the estimation of three main effects (i.e., *x*, *y*, and *z* in this example), three 2-way interactions (i.e., *xy*, *xz*, and *yz*), and one 3-way interaction (i.e., *xyz*). Table IV-5 shows the design matrix for the airline passenger screening systems experiment that includes the interactions of the main effects. The sign values for the interaction columns are determined by multiplying the signs of the factors of interest. Note that for the main effects, the design moves one factor from its low to its high setting while holding the other factors constant.

Table IV-5. 2 x 3 study design matrix.

Design point	*x*	*y*	*z*	*xy*	*xz*	*yz*	*xyz*	Response
DP_1	–1	–1	–1	+1	+1	+1	–1	R_1
DP_2	+1	–1	–1	–1	–1	+1	+1	R_2
DP_3	–1	+1	–1	–1	+1	–1	+1	R_3
DP_4	+1	+1	–1	+1	–1	–1	–1	R_4
DP_5	–1	–1	+1	+1	–1	–1	+1	R_5
DP_6	+1	–1	+1	–1	+1	–1	–1	R_6
DP_7	–1	+1	+1	–1	–1	+1	–1	R_7
DP_8	+1	+1	+1	+1	+1	+1	+1	R_8

The HSI consultant runs the 2 x 3 study design matrix as just discussed using eight judiciously chosen subjects. Before proceeding further, however, a word of caution is warranted with regards to the choice of said subjects. Such a study as described here lacks a key ingredient of classic design of experiments—random assignment. As a result, the study design is really a quasi-experimental study design, which raises obvious potential problems of internal validity. The biggest threat to internal validity is selection—that the subjects differ in other important ways besides the factors considered by the experimenter. Additionally, since each subject individually represents a prespecified design point, it is imperative that they actually embody the corresponding combination of factor levels. Consequently, one needs to be cognizant in opting for such a study of the implicit tradeoff between risk for threats to validity and economy of effort, both in terms of time and resources.

Mindful of these issues, the HSI consultant is duly vigilant for potential threats when selecting subjects and conducting the experiment. The experiment yields the results shown in Table IV-6 where the response (R) is the calculated number of missed threats per thousand passengers screened.

Table IV-6. Initial results for the airline passenger screening system experiment.

Design point	*x*	*y*	*z*	*xy*	*xz*	*yz*	*xyz*	Response
DP_1	–1	–1	–1	+1	+1	+1	–1	8.0
DP_2	+1	–1	–1	–1	–1	+1	+1	2.3
DP_3	–1	+1	–1	–1	+1	–1	+1	5.7
DP_4	+1	+1	–1	+1	–1	–1	–1	0.2
DP_5	–1	–1	+1	+1	–1	–1	+1	9.0
DP_6	+1	–1	+1	–1	+1	–1	–1	3.0
DP_7	–1	+1	+1	–1	–1	+1	–1	6.9
DP_8	+1	+1	+1	+1	+1	+1	+1	1.5

The next step is to calculate effects sizes for each of the factors and the interactions between the factors. As previously mentioned, the effect size is the average change in the response resulting from moving a factor from its (–1) to its (+1) level while holding all other factors fixed. Effect size is also a convenient means for gauging the relative importance of factors. A simple shortcut for calculating effect sizes is to apply the signs of the factor or interaction column to the corresponding response, sum them, and then divide by 2^{k-1}, where k is the number of factors. Effect size calculations for the three main effects, three 2-way interactions, and single 3-way interaction in the airline passenger screening systems experiment are shown below (Equations 43–49).

$$\begin{aligned} e_x &= \frac{-R_1 + R_2 - R_3 + R_4 - R_5 + R_6 - R_7 + R_8}{4} \\ &= \frac{-8.0 + 2.3 - 5.7 + 0.2 - 9.0 + 3.0 - 6.9 + 1.5}{4} = -5.68 \end{aligned} \tag{43}$$

$$\begin{aligned} e_y &= \frac{-R_1 - R_2 + R_3 + R_4 - R_5 - R_6 + R_7 + R_8}{4} \\ &= \frac{-8.0 - 2.3 + 5.7 + 0.2 - 9.0 - 3.0 + 6.9 + 1.5}{4} = -1.98 \end{aligned} \tag{44}$$

$$e_z = \frac{R_1 + R_2 + R_3 + R_4 - R_5 - R_6 - R_7 - R_8}{4}$$
$$= \frac{8.0 + 2.3 + 5.7 + 0.2 - 9.0 - 3.0 - 6.9 - 1.5}{4} = 1.08 \tag{45}$$

$$e_{xy} = \frac{R_1 - R_2 - R_3 + R_4 + R_5 - R_6 - R_7 + R_8}{4}$$
$$= \frac{8.0 - 2.3 - 5.7 + 0.2 + 9.0 - 3.0 - 6.9 + 1.5}{4} = 0.20 \tag{46}$$

$$e_{xz} = \frac{R_1 - R_2 + R_3 - R_4 - R_5 + R_6 - R_7 + R_8}{4}$$
$$= \frac{8.0 - 2.3 + 5.7 - 0.2 - 9.0 + 3.0 - 6.9 + 1.5}{4} = 0.20 \tag{47}$$

$$e_{yz} = \frac{R_1 + R_2 - R_3 - R_4 - R_5 - R_6 + R_7 + R_8}{4}$$
$$= \frac{8.0 + 2.3 - 5.7 - 0.2 - 9.0 - 3.0 + 6.9 + 1.5}{4} = 0.20 \tag{48}$$

$$e_{xyz} = \frac{-R_1 + R_2 + R_3 - R_4 + R_5 - R_6 - R_7 + R_8}{4}$$
$$= \frac{-8.0 + 2.3 + 5.7 - 0.2 + 9.0 - 3.0 - 6.9 + 1.5}{4} = 0.10 \tag{49}$$

At this point, the HSI consultant has accomplished the first step in Simon's research strategy, namely, conducting a screening factorial study. The next step is to conduct a Pareto analysis to reduce the complexity of the solution space by eliminating those factors and interactions between factors that fail to make a meaningful contribution. Figure IV-24 displays the results of the full factorial airline passenger screening systems experiment.

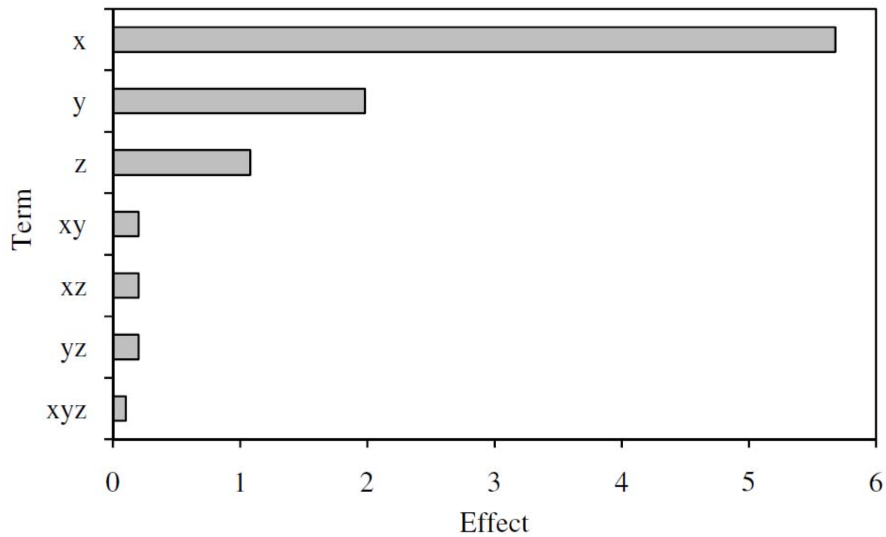


Figure IV-24. Pareto analysis of the airline passenger screening system experiment.

It is evident from Figure IV-24 that individual aptitude (*x*) is the most important factor in explaining performance, followed by training time (*y*) and then equipment (*z*). This result confirms the HSI consultant's original concern that user attributes could impact reported system reliability. It also appears that the HSI consultant can safely disregard the interaction terms and focus just on the factors, thereby simplifying subsequent calculations.

5. Defining a Tradeoff Function

To quantitatively consider tradeoffs requires the formulation of some functional model relating a performance of interest to its determinants. In the case of the airline passenger screening system experiment, it is possible to formulate a linear model directly from the study results in Table IV-6. The model will be of the following general form:

$$R = ax + by + cz + d \tag{50}$$

where R is the response, x is ±1 depending on choice of personnel domain factor level (i.e., aptitude score of 60 or 100), y is ±1 depending on choice of training domain factor level (i.e., training time of 10 or 50 hours), and z is ±1 depending on choice of equipment. The model parameters for each factor (i.e., a, b, and c) can be calculated in a manner that is similar to that used for calculating effect size. Again, we apply the signs of the factor

column to the corresponding response (R), sum them, and this time divide by 2^k, where k is the number of factors. For example, to calculate the model parameter associated with individual aptitude (a):

$$\begin{aligned} a &= \frac{-R_1 + R_2 - R_3 + R_4 - R_5 + R_6 - R_7 + R_8}{8} \\ &= \frac{-8.0 + 2.3 - 5.7 + 0.2 - 9.0 + 3.0 - 6.9 + 1.5}{8} = -2.84 \end{aligned} \tag{51}$$

The values for the parameters b and c are similarly calculated:

$$\begin{aligned} b &= \frac{-R_1 - R_2 + R_3 + R_4 - R_5 - R_6 + R_7 + R_8}{8} \\ &= \frac{-8.0 - 2.3 + 5.7 + 0.2 - 9.0 - 3.0 + 6.9 + 1.5}{8} = -0.99 \end{aligned} \tag{52}$$

$$\begin{aligned} c &= \frac{R_1 + R_2 + R_3 + R_4 - R_5 - R_6 - R_7 - R_8}{8} \\ &= \frac{8.0 + 2.3 + 5.7 + 0.2 - 9.0 - 3.0 - 6.9 - 1.5}{8} = 0.54 \end{aligned} \tag{53}$$

The value of the parameter d is simply the average response. That is:

$$\begin{aligned} d &= \frac{R_1 + R_2 + R_3 + R_4 + R_5 + R_6 + R_7 + R_8}{8} \\ &= \frac{8.0 + 2.3 + 5.7 + 0.2 + 9.0 + 3.0 + 6.9 + 1.5}{8} = 4.57 \end{aligned} \tag{54}$$

Now the HSI consultant can calculate an estimated response for any combination of the three factors at either of their two levels using the following linear regression model:

$$R = -2.84x - 0.99y + 0.54z + 4.57 \tag{55}$$

Hence, the HSI consultant can provide the project manager with an estimate of how well a system comprised of a TSA employee with an aptitude score of 60 (x = –1) and 50 hours of training (y = +1) using a Magic Eyes scanner (z = –1) will perform:

$$R = -2.84(-1) - 0.99(+1) + 0.54(-1) + 4.57 = 5.8 \tag{56}$$

But what about an employee with an aptitude score of 80 and 40 hours of training? To answer these types of questions, the values of aptitude score and training must be scaled so they vary between –1 and +1 over the range between the low and high factor level settings. For example, consider the following scaling for aptitude score:

$$x = \frac{score - 80}{20} \tag{57}$$

A score of 60 corresponds to $x = -1$ and a score of 100 corresponds to $x = +1$, which matches the low and high factor level settings respectively. Similarly, an appropriate scaling for training would be:

$$y = \frac{training - 30}{20} \tag{58}$$

There is no need to scale the equipment factor as it is only defined for the values $z = \pm 1$. Substituting the scaled terms for x and y into Equation 55 along with the substitution $z = equipment$, yields:

$$R = a\left(\frac{score - 80}{20}\right) + b\left(\frac{training - 30}{20}\right) + c\left(equipment\right) + d \tag{59}$$

where a, b, c, and d are the same model parameters calculated earlier. Substituting the estimates for these model parameters into Equation 59 and using some simple algebra yields:

$$R = -0.14\left(score\right) - 0.05\left(training\right) + 0.54\left(equipment\right) + 17.41 \tag{60}$$

Equation 60 is a *tradeoff function* in terms of the personnel domain (aptitude score), training domain (training time), and human factors engineering domain (equipment) as they relate to system performance. Using Equation 60, the HSI consultant can estimate system performance as a function of these three HSI domains for any point in the shaded areas of the solution space shown in Figure IV-25.

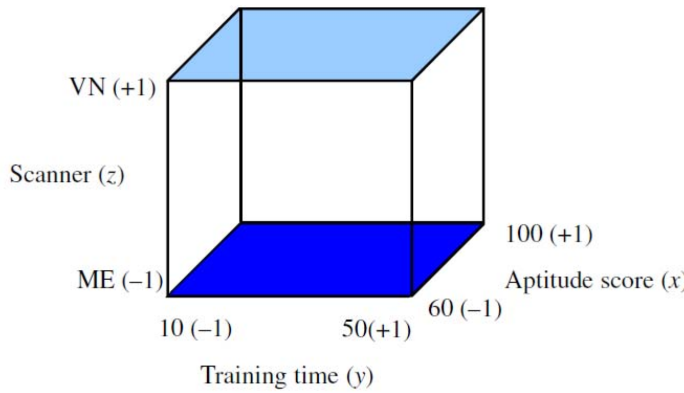


Figure IV-25. Potential solution space, shown as shaded areas, for the airline passenger screening system experiment.

6. Developing Isoperformance Curves

The idea of developing isoperformance curves among HSI domains was first advanced as a tradeoff methodology by Kennedy and Jones. The central reasoning behind their isoperformance methodology is the idea that, almost always, a specified level of performance can be produced by more than one combination of factors. Hence, isoperformance curves trace all combinations of two or more HSI domains that produce a specified level of performance. Figure IV-26 depicts a hypothetical pair of isoperformance curves that might be produced by an experiment such as the one just performed to evaluate the airline passenger screening systems. On the graph, training time is plotted on the abscissa and personnel aptitude score is plotted on the ordinate. The two lines correspond to combinations of personnel aptitude, training time, and equipment that produce equivalent performance, for example, 3 missed threats per thousand passengers screened. Hence, performance at points R1, R2, and R3 is the same, but the contributions from various combinations formed by personnel, training, and equipment factors are different. All sorts of tradeoffs can be explored using these plots. As an example, for equipment A, if the aptitude of newly hired screeners is decreased from P2 to P1, then we will need to increase training from T1 to T2 to maintain system performance. Alternatively, we could elect to adopt equipment B and continue to train only to T1. This is exactly the type of information decision makers need to consider various courses of action.

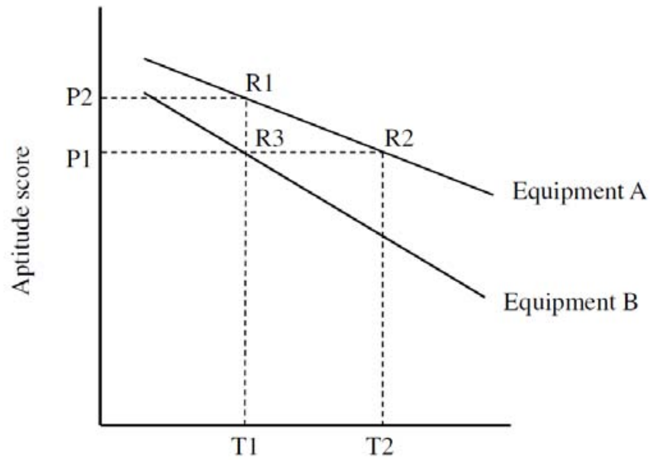


Figure IV-26. Hypothetical isoperformance plots for two equipment configurations.

Isoperformance curves provide equivalent options in terms of the factors included in the underlying model. However, other external factors such as cost, risk, or utility must then be used to make a final selection from among the range of equivalent solutions. For example, Figure IV-27 depicts the fact that there is an upper limit on the aptitude scores of potential new screeners. Likewise, there is an upper limit on the amount of training that can be provided. Given these constraints, we see that there are no feasible solutions involving equipment A. Additionally, only a third of the solutions on the isoperformance curve for equipment B are feasible.

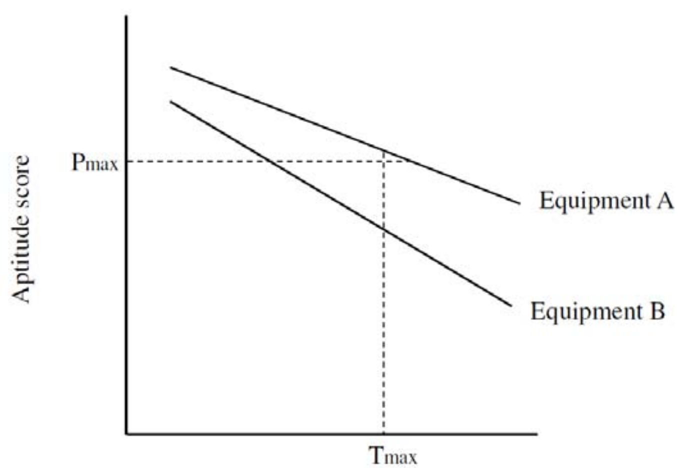


Figure IV-27. Hypothetical isoperformance plots for two equipment configurations with personnel and training domain feasibility constraints.

To create isoperformance curves for the airline passenger screening systems problem, it is necessary to first establish a target level of performance, R_{iso}. In so doing, the value of R in the linear regression model given by Equation 60 becomes fixed at R_{iso}:

$$R_{iso} = -0.14(score) - 0.05(training) + 0.54(equipment) + 17.41 \tag{61}$$

Recalling that *equipment* = 1 for Virtually Nude and –1 for Magic Eyes, two isoperformance equations can be formulated corresponding to the two equipment options:

Virtually Nude:

$$\begin{aligned} R_{iso} &= -0.14(score) - 0.05(training) + 0.54(+1) + 17.41 \\ &= -0.14(score) - 0.05(training) + 17.95 \end{aligned} \tag{62}$$

Magic Eyes:

$$\begin{aligned} R_{iso} &= -0.14(score) - 0.05(training) + 0.54(-1) + 17.41 \\ &= -0.14(score) - 0.05(training) + 16.87 \end{aligned} \tag{63}$$

By simply rearranging terms, *score* can be expressed in terms of a simple linear function of *training* for each of the equipment options:

Virtually Nude:

$$score = -0.36(training) - 7.14R_{iso} + 128.21 \tag{64}$$

Magic Eyes:

$$score = -0.36(training) - 7.14R_{iso} + 120.50 \tag{65}$$

For those who find that these equations are not intuitive, it may help to recognize that they reduce to the following general form:

$$score = m(training) + n \tag{66}$$

which is the formula for a line of slope m and intercept n. It is also worth noting here that, given a specified level of performance (R_{iso}), each of these equations describes equivalent combinations of the determinants, individual aptitude and training time, within one of the two planes comprising the solution space (Figure IV-28).

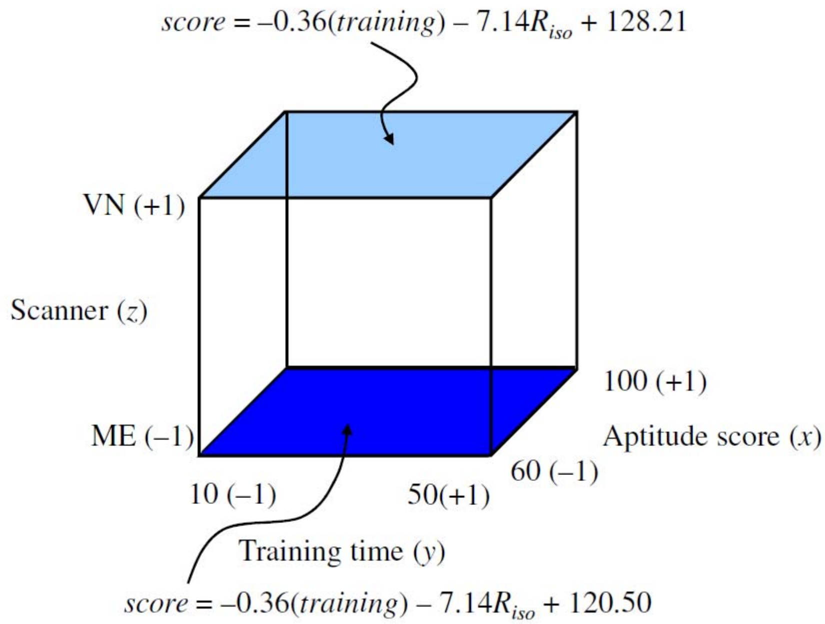


Figure IV-28. Potential solution space, shown as shaded areas, for the airline passenger screening system experiment with corresponding isoperformance equations.

By picking several representative values for training (e.g., 6, 12, 24, 30, 36, 42, 48, and 54 hours) and using Equations 64 and 65, the HSI consultant can compute corresponding aptitude scores that will yield a performance of R_{iso} for each equipment configuration. The HSI consultant decides to create isoperformance curves for several performance levels: high (R_{iso} = 1), moderate (R_{iso} = 3) and low (R_{iso} = 5). These performance levels correspond to system reliabilities of 0.999, 0.997, and 0.995 respectively. The resulting isoperformance readouts for each of the equipment options are shown in Figure IV-29.

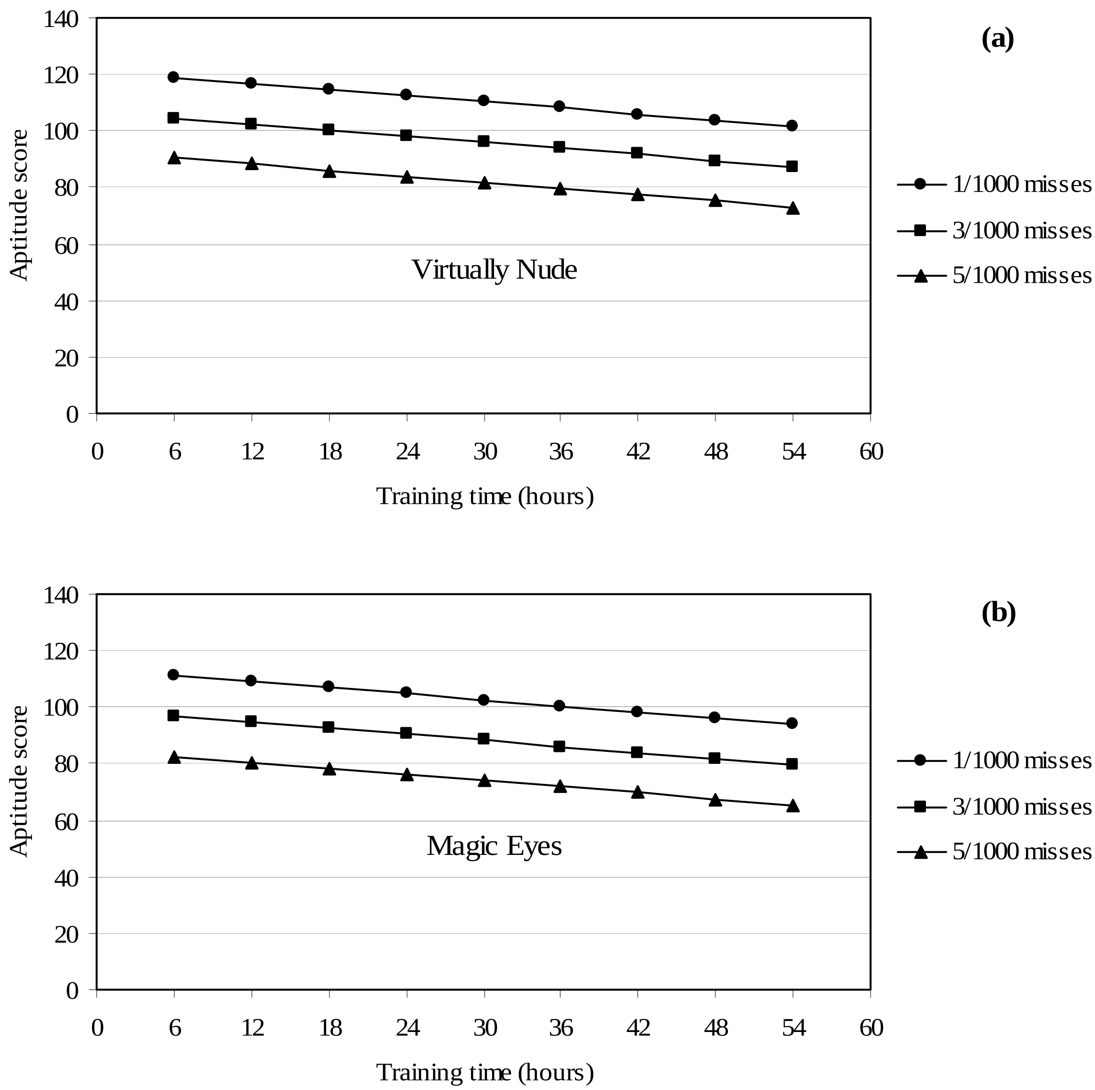


Figure IV-29. Isoperformance curves for the two equipment options: a) Virtually Nude, b) Magic Eyes.

The relative flat slope of the isoperformance curves suggest that performance is more sensitive to aptitude than training—so personnel domain considerations regarding the range of aptitudes to accommodate need to be taken into account during the systems decision process. Given the aptitude and training constraints provided by the human resources director, it is readily evident that the reliability for Virtually Nude (Figure IV-29a) reported in the RFP (i.e., one missed threat per thousand passengers screened) is not obtainable, even when we consider a TSA employee of average aptitude. If we consider

the suggested threshold-level personnel and training domain requirements, which you will recall was an aptitude score of 80 and a training time of 40 weeks, the best reliability that can be achieved with Virtually Nude is five missed threats per thousand passengers screened. In the case of Magic Eyes (Figure IV-29b), a reliability of three missed threats per thousand passengers screened, as was claimed in the RFP, is achievable, but only with the maximum acceptable length of training. Again, given the suggested threshold-level personnel and training domain requirements, the best reliability that can be achieved is approximately four missed threats per thousand passengers screened. As this quick analysis demonstrates, quite a bit of useful information can be extracted from graphical isoperformance readouts of the type illustrated in Figure IV-29.

7. Designing a System Solution

It is finally possible to look at complete system solutions—complete in the sense that information is available about the liveware, hardware, and software components of the system and how their attributes contribute to overall system performance. However, in considering the information provided by the isoperformance curves, the project team is faced with potentially conflicting objectives. For example, the objective to maximize reliability would guide decision makers to select a system based on the performance achievable given the maximum feasible training time (i.e., 54 hours) and with minimal accommodation of the range of potential user aptitudes (i.e., assuming a minimum aptitude score of 80). In contrast, the human resource director's objectives would guide decision makers to accept a lower level of system reliability in exchange for minimizing training time and maximally accommodating the range of potential users—both considerations with significant implications for the cost of ownership for a system. How then should the conflicting objectives in this decision problem be resolved?

At this point in the airline passenger screening systems problem, it is necessary to directly address the issue raised by Kennedy and Jones of coupling isoperformance and utility analysis—what is the acceptable level of performance for which we need to consider some combinations of equipment, user aptitude, and training? In this problem context, the performance of interest is reliability and it can be described in terms of an

isoperformance model involving equipment, aptitude, and training. The achieved level of reliability contributes to the overall utility of the solution, but aptitude and training do as well and not necessarily in a convergent manner. This then brings us to the subject of multiple criteria decision making, where it is was shown in Section IV-F that physical programming provides an attractive methodology for integrating the isoperformance approach within an overarching utility analysis—and without requiring decision makers to undertake the dubious task of explicitly developing subjective weight schemes.

To make use of a physical program, we first need to reformulate Equation 61 so that z is redefined as a binary decision variable, taking on a value of 1 for Virtually Nude and 0 for Magic Eyes. This is a slight change from the coding used previously in the design of experiments and is mainly for the convenience of modeling as most optimization software utilize binary indicator variables. Accordingly, in the airline passenger screening systems source selection decision, the criteria of interest to the decision maker are as follows:

Aptitude:

$$g_1 \equiv y \tag{67}$$

Training:

$$g_2 \equiv x \tag{68}$$

Reliability:

$$g_3 \equiv r_{iso} = -0.14x - 0.05y + 1.08z + 16.87 \tag{69}$$

Capacity:

$$g_4 \equiv c_0 \left(1 - z\right) + c_1 z \tag{70}$$

Privacy:

$$g_5 \equiv p_0 \left(1 - z\right) + p_1 z \tag{71}$$

Manpower:

$$g_6 \equiv m_0 \left(1 - z\right) + m_1 z \tag{72}$$

where x is training time in hours; y is aptitude score; z is choice of system; r_{iso} is estimated reliability in missed threats per thousand passengers screened; c_z is the

capacity in number of passengers per hour of system $z = 0,1$; p_z is the privacy of system $z = 0,1$; and m_z is the number of operators needed for system $z = 0,1$. The decision variables are

$$\psi = \{x, y, z\}$$

The various wishes and specifications expressed by the systems decision stakeholders are as follows:

- Reliability: In terms of the number of missed targets per thousand passengers screened, lower is better. The rate of missed targets must be less than 7.5, representing a 25% improvement over the current system—otherwise there is no point to the acquisition.
- Aptitude: The lower the score the better as the pool of new hires that can be easily accommodated is increased.
- Training: Shorter is better to decrease resource utilization and ownership costs.
- Capacity: Higher is better to decrease the impact on airline passengers and airline operators.
- Privacy: In terms of the constructive scores, higher is better to ease concerns about the potential impact on passengers' civil liberties.
- Manpower: Less is better, again to decrease ownership costs.

The choice of class functions and the range limits that delineate degrees of desirability as expressed by the stakeholders are reported in Table IV-7.

Table IV-7. Physical programming region limits table.

i^{th} criteria	Class type	t_{i1}^*	t_{i2}^*	t_{i3}^*	t_{i4}^*	t_{i5}^*
Aptitude	1-S	60	70	75	78	80
Training	1-S	20	30	35	40	54
Reliability	1-S	1	2	3	5	7.5
Capacity	2-S	120	95	70	45	20
Privacy	2-S	1	0	-1	-2	-3
Manpower	1-S	0	1	2	3	4

$t_{is}^* = t_{is}^+$ for class 1-S functions, t_{is}^- for class 2-S functions

The physical programming problem statement takes the following form:

$$\min_{d_{is}^-,\, d_{is}^+,\, x,\, y,\, z} \quad J = \sum_{i=1}^{n_{sc}} \sum_{s=2}^{5} \left(\tilde{w}_{is}^- d_{is}^- + \tilde{w}_{is}^+ d_{is}^+ \right) \tag{73}$$

subject to

$$\left.\begin{array}{l} g_i - d_{is}^+ \le t_{i(s-1)}^+ \\ g_i \le t_{i5}^+ \\ d_{is}^+ \ge 0 \end{array}\right. \quad \left(\text{for all } i \text{ in class 1S, } i = \{1,2,3,6\},\ s = 2,\ldots,5\right)$$

$$\left.\begin{array}{l} g_i + d_{is}^- \ge t_{i(s-1)}^- \\ g_i \ge t_{i5}^- \\ d_{is}^- \ge 0 \end{array}\right. \quad \left(\text{for all } i \text{ in class 2S, } i = \{4,5\},\ s = 2,\ldots,5\right)$$

$$\begin{array}{l} x \ge 0 \\ y \ge 0 \\ 0 \le z \le 1 \end{array}$$

We relaxed the binary constraint on z in Equation 73 so that the program can be solved as a linear physical program with the understanding that this approach is tenable only if the optimal solution sets z to either 0 or 1 (which it does). Otherwise, Equation 73

would need to be solved as an integer linear program. The weights generated using the physical programming weight algorithm and the optimal decision results are presented in Table IV-8.

Table IV-8. Physical programming optimization results.

i^{th} criteria	Weights				Value
	$\tilde{w}_{i2}^{+}$	$\tilde{w}_{i3}^{+}$	$\tilde{w}_{i4}^{+}$	$\tilde{w}_{i5}^{+}$	
Aptitude	0.03	0.25	1.24	1.46	70
Training	0.03	0.25	2.25	18.28	41.9
Reliability	0.25	1.13	2.41	12.86	7.5
Capacity	0.01	0.05	0.25	1.36	60
Privacy	0.25	1.13	6.19	34.03	-1
Manpower	0.25	1.13	6.19	34.03	2

The optimal setting of the decision variables is $\psi^* = \{41.9, 70, 0\}$, which corresponds to selecting the Magic Eyes system, providing approximately 42 hours of training, and accommodating employees with a minimum aptitude score of 70. It is worth noting that the most utility is attained by accepting the minimally tolerable performance increment in reliability (i.e., a miss rate of 7.5 targets per thousand passengers screened) rather than maximizing reliability at the expense of the personnel and training domains. Moreover, the optimal system performance, given consideration of the HSI dimensions of the solution space, differs by a factor of 2.5 from that reported in the RFP. Thus, this illustrates the importance of giving due regard to HSI considerations early in the systems decision process!

It should also be noted that the reliability level attained through physical programming differs from what would have been anticipated had we only used the isoperformance readouts and considered the HSI domain issues solely as problem constraints. For example, strictly considering the isoperformance readout for Magic Eyes (Figure IV-29) would have led to the conclusion that the reported reliability in the RFP of

three missed threats per thousand passengers screened is both reasonable and feasible. However, the necessary training time would have been highly undesirable from the perspective of the human resources department. Likewise, personnel considerations would have been limited to accommodating only minus two standard deviations—that is, 97.5 percent of employees. Had we allowed performance considerations to dominate the systems decision process, the end result might have been an unbalanced solution that maximized operational effectiveness (at least, in the short term) at the expense of operational suitability. However, the physical programming problem statement, by including HSI considerations directly in the problem objective function, suggested a more balanced solution that considered both operational effectiveness and suitability as independent dimensions of the solution space. This observation reinforces the importance of coupling isoperformance with utility analysis as suggested by Kennedy and Jones.

H. CONCLUSION

A complete decision analysis should include, at a minimum, both a sensitivity analysis and a cost/benefit analysis. Nonetheless, we will stop at this point in the airline passenger screening systems problem as the primary purpose of the example was to convey a more concrete, and therefore mentally tractable, image of the HSI trade space in contrast to the abstract perspective provided by the conceptual models introduced earlier in the chapter. The example illustrates how we can take a significant step forward in deliberatively considering the HSI trade space in the systems decision process by integrating Simon's research strategy, Kennedy and Jones' isoperformance approach, and coupling isoperformance with utility analysis through such means as physical programming. Hopefully, experienced HSI practitioners will take a moment to pause and reflect that, despite this being a simple "toy" problem, the choice of an isoperformance criterion that balanced operational effectiveness and suitability considerations was not intuitively self-evident. Consequently, a systematic and disciplined process is needed when approaching the problem of planning and designing complex systems with desired outcomes.

With that said, we have only started the analysis of what DePuy and Bonder (1982) call the micro MPT supply and demand interface—or "little HSI" as my fellow HSI practitioner, John Burns, described it. What should be self evident is that attending to the macro MPT supply and demand interface—or "big HSI" in Burn's lexicon—will involve massive amounts of data. Managing this interface can only be made tractable through the mathematical tools and methods of operations research, and probably, the incorporation of some elements of computational design at the level of little HSI. Thus, it is promising that our little HSI process culminates in an operations research method—physical programming—by which we can include considerations from the big HSI process in the form of Messac and colleagues' (1996) hard type of constraints.

From a more philosophical perspective, another outstanding issue that needs to be addressed is how we incorporate this macro/micro or big/little duality into our conceptualization of HSI. Little HSI concentrates on individual technological systems and subsystems and, at least in its contemporary implementation, is strongly oriented towards human factors engineering (see Pew & Mavor, 2007), or what Meister (1999) terms "microergonomic," considerations. In contrast, big HSI focuses on the development and utilization of human resources within organizations that own and operate technological systems that are, in turn, the subject matter of little HSI; it is concerned mainly with macroergonomic considerations of organizational and work-system design. While little HSI pursues local optima for individual systems, big HSI seeks a global optimum across systems. Here, then, is the potential for little HSI to work at cross purposes with big HSI, because local optima do not always pave the way to a global optimum. Accordingly, the overarching goal of HSI must be one of making organizationally net positive contributions, otherwise we risk creating solutions today that are tomorrow's problems.

I. BIBLIOGRAPHY

Bar-Yam, Y. (2003). When systems engineering fails: towards complex systems engineering. In *IEEE International Conference on Systems, Man and Cybernetics* (Vol. 2, pp. 2021–2028). Piscataway, NJ: IEEE Press.

Barnes, M., & Beevis, D. (2003). Human systems measurements and trade-offs in system design. In H. R. Booher (Ed.), *Handbook of human systems integration* (chap. 8, pp. 233−263). New York, NY: John Wiley and Sons.

Blanchard, B. S., & Fabrycky, W. J. (2006). *Systems engineering and analysis* (4^{th} ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Booher, H. R. (Ed.). (1990). *MANPRINT: an approach to systems integration*. New York, NY: Van Nostrand Reinhold.

Booher, H. R. (2003). Introduction: human systems integration. In H. R. Booher (Ed.), *Handbook of human systems integration* (Chap. 1, pp. 1−30). New York, NY: John Wiley and Sons.

Booher, H. R., Beaton, R. J., & Greene, F. (2009). Human systems integration. In. A. P. Sage & W. B. Rouse (Eds.), *Handbook of systems engineering and management* (Chap. 31, pp. 1319−1359). Hoboken, NJ: John Wiley and Sons.

Cagle, D. [Drawing of semi-naked Statue of Liberty] (2010, January 14). *Daryl Cagle's Cartoon Web Log*. Retrieved March 13, 2010, from http://blog.cagle.com/daryl /2010/01/14/yes-i-look-for-any-excuse-to-draw-the-statue-of-liberty-naked/

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: a handbook for research on interactions*. New York, NY: Irvington Publishers.

Department of the Army (1976). *Personnel-material systems: human factors engineering program* (AR 602-1). Washington, DC: Author.

Department of Defense (1991). *Defense acquisition management policies and procedures* (DoDI 5000.2). Washington, DC: Author.

DePuy, W. E., & Bonder, S. (1982). *Integration of MPT supply and demand and the system acquisition process* (Research Note No. 82-16). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

de Weck, O. L., & Jones, M. B. (2006). Isoperformance: analysis and design of complex systems with desired outcomes. *Systems Engineering*, 9(1), 45−61.

Eigen, M., & Winkler, R. (1981). *Laws of the game: how the principles of nature govern chance* (R. Kimber & R. Kimber, Trans.). New York, NY: Alfred A. Knopf, Inc. (Original work published 1965).

Hitchins, D. K. (1992). *Putting systems to work.* Chichester, England: John Wiley and Sons.

Horne, D. K. (1986). *The impact of soldier quality on performance in the Army* (Report No. ARI-TR-708). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Jones, M. B. (2000). Isoperformance and personnel decisions. *Human Factors*, 42(2), 299−317.

Jones, M. B., & Kennedy, R. S. (1996). Isoperformance curves in applied physiology. *Human Factors*, 38(1), 167−182.

Jones, M. D., Kennedy, R. S., Baltzley, D. R., & Westra, D. P. (1988). *Omega square analyses and reevaluation of a meta-analysis for equipment features and individual differences in an aircraft simulator*. Orlando, FL: Essex Corporation.

Jones, M. D., Kennedy, R. S, Kuntz, L. A., & Baltzey, D. R. (1987). Isoperformance: trading off selection, training, and equipment variations to maintain the same level of systems performance. In *Proceedings of the Human Factors Society 31st Annual Meeting 1987* (pp. 634−637). Santa Monica, CA: Human Factors Society.

Jones, M. B., Kennedy, R. S., & Stanney, K. M. (2004). Toward systematic control of cybersickness. *Presence*, 13, 589-600.

Jones, M. B., Kennedy, R. S., Turnage, J. J., Kuntz, L. A., & Jones, S. A. (1985). *Meta-analysis of human factors engineering studies comparing individual differences, practice effects and equipment design variations* (Report No. EOTR-86-7). Wright-Patterson Air Force Base, OH: Aeronautical Systems Division/PMRSB. (DTIC Accession No. ADA168790)

Jones, M. B., & Turnage, J. J., & Kennedy, R. S. (1993). *Isoperformance as a framework for human systems integration* (Report No. HSC-TR-93-001). Brooks Air Force Base, TX: Human Systems Center.

Kennedy, B. S., & Jones, M. B. (1992). *Simulating the impact of MPTS trade-off decisions by application of the isoperformance methodology*. Orlando, FL: Naval Training Systems Center. (DTIC Accession No. ADA274607)

Kennedy, R. S., Jones, M. B., & Baltzley, D. R. (1988). *Empirical demonstration of isoperformance methodology preparatory of an interactive expert computerized decision aid* (Report No. ARI Research Note 88-93. Alexandria, VA: United States Army Research Institute for the Behavioral and Social Sciences. (DTIC Accession No. ADA202429)

Kennedy, R. S., Jones, M. B., & Baltzley, D. R. (1989). *Regions of isoperformance: an interactive software package for trading off training, personnel, and equipment* (Report No. AAMRL-TR-89-003). Wright-Patterson Air Force Base, OH: Harry G. Armstrong Aerospace Medical Research Laboratory. (DTIC Accession No. ADB136770)

Kennedy, R. S., Jones, M .B., & Turnage, J. J. (1990, October). *A meta-model for systems development through life cycle phases: coupling the isoperformance methodology with utility analysis*. Presented at the Aerospace Technology Conference and Exposition, Long Beach, CA.

Lintern, G., Nelson, B. E., Sheppard, D. J., Westra, D. P., & Kennedy, R. S. (1981). *Visual technology research simulator (VRTS) human performance research: phase III* (Report No. NAVTRAEQUIPCEN 78-C-0060-11). Orlando, FL: Naval Training Equipment Center. (DTIC Accession No. ADA112475)

Matthews, G., Davies, D. R., Westerman, S. J., & Stammers, R. B. (2000). *Human performance: cognition, stress, and individual differences*. Philadelphia, PA: Taylor and Francis.

McCauley, M. E., & Kennedy, R. S. (1976). *Human exposure limits for very low frequency vibration* (Report No. TP-76-36). Point Mugu, CA: U.S. Navy Pacific Missile Test Center.

Meister, D. (1999). *The history of human factors and ergonomics*. Mahwah, NJ: Lawrence Erlbaum Associates.

Messac, A. (1996). Physical programming: effective optimization for computational design. *AIAA Journal*, 34(1), 149–158.

Messac, A., Gupta, S. M., Akbulut, B. Linear physical programming: a new approach to multiple objective optimization. *Transactions on Operational Research*, 8, 39–59.

Miller, N. L., & Firehammer, R. (2007). Avoiding a second hollow force: the case for including crew endurance factors in the afloat staffing policies of the U.S. Navy. *Naval Engineers Journal*, 1, 83–96.

Miller, N. L., & Shattuck, L. G. (2007, January). *A re-conceptualization of human systems integration in theory and practice*. Presentation at the International Council of Systems Engineering International Workshop, Albuquerque, New Mexico.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Parnell, G. S., Driscoll, P. J., Henderson, D.L. (Eds.) (2008). *Decision making in systems engineering and management*. Hoboken, NJ: John Wiley and Sons.

Pasmore, W. A., & Sherwood, J. J. (1978). Organizations as sociotechnical systems. In W. A. Pasmore & J. J. Sherwood (Eds.), *Sociotechnical systems: a source book* (pp. 3–7). La Jolla, CA: University Associates.

Pew, R. W., & Mavor, A. S. (2007). *Human-system integration in the system development process: a new look*. Washington, DC: National Academies Press.

Simon, C. W. (1970). *The use of central-composite designs in human factors engineering experiments* (Report No. AFOSR-TR-72-1637). Arlington, VA: Air Force Office of Scientific Research.

Simon, C. W. (1973). *Economical multifactor designs for human factors engineering experiments* (Report No. AFOSR-TR-73-1702). Arlington, VA: Air Force Office of Scientific Research. (DTIC Accession No. AD767739)

Simon, C. W. (1974). *Methods for handling sequence effects in human factors engineering experiments* (Report No. HAC-P74-54). Arlington, VA: Air Force Office of Scientific Research. (DTIC Accession No. ADA035109)

Simon, C. W. (1975). *Methods for improving information from "undesigned" human factors experiments* (Report No. HAC-P75-287). Arlington, VA: Air Force Office of Scientific Research. (DTIC Accession No. ADA018455)

Simon, C. W. (1976a). *Analysis of human factors engineering experiments: characteristics, result, and applications* (Report No. AFOSR-TR-77-0333). Washington, DC: Air Force Office of Scientific Research. (DTIC Accession No. ADA038184)

Simon, C. W. (1976b). *Response surface methodology revisited: a commentary on research strategy* (Report No. CWS-01-76). Washington, DC: Air Force Office of Scientific Research.

Simon, C. W. (1977). *New research paradigm for applied experimental psychology: a system approach* (Report No. CWS-04-77). Washington, DC: Air Force Office of Scientific Research.

Simon, C. W. (1978). *Applications of advanced experimental methodologies to AWAVS training research* (Report No. NAVTRAEQUIPCEN 77-C-0065-1). Orlando, FL: Naval Training Equipment Center.

Simon, C. W. (1984). *Handling bias from individual differences in between-subject holistic experimental designs* (Report No. NAVTRASYSCEN 81-C-0105-10). Orlando, FL: Naval Training Systems Center.

Simon, C. W. (1985). *Handling bias from individual differences in between-subject holistic experimental designs* (Report No. NAVTRASYSCEN 81-C-0105-10). Orlando, FL: Naval Training Systems Center.

Simon, C. W. (1987). *Investigation into the use of normal and half-normal plots for interpreting results from screening experiments* (Report No. NAVTRASYSCEN 85-D-0026-4). Orlando, FL: Naval Training Systems Center.

Simon, C. W., & Roscoe, S. N. (1981). *Application of a multifactor approach to transfer of training research* (Report No. NAVTRAEQUIPCEN 78-C-0060-6). Orlando, FL: Naval Training Equipment Center.

Simon, H. A. (1996). *The sciences of the artificial* (3^{rd} ed.). Cambridge, MA: MIT Press.

Snow, R. E. (1978). Theory and method for research on aptitude processes. *Intelligence,* 2, 225−278.

Snow, R. E. (1989). Aptitude-treatment interaction as a framework for research on individual differences in learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences: advances in theory and research* (Chap. 2, pp. 13−59). New York, NY: W. H. Freeman and Company.

Turnage, J. J., Kennedy, R. S., & Jones, M. B. (1990, November). Trading off performance, training, and equipment factors to achieve similar performance. *Proceedings of the 32nd Annual Conference of the Military Testing Association*. Orange Beach, AL: Naval Education & Training Program Management Support Activity.

West, P. D. (2008). Solution design. In G. S. Parnell, P. J. Driscoll, & D. L. Henderson, (Eds.) (2008). *Decision making in systems engineering and management* (Chap 11, pp. 317−355). Hoboken, NJ: John Wiley and Sons.

Whitener, E. M. (1989). A meta-analytic review of the effect on learning of the interaction between prior achievement and instructional support. *Review of Educational Research*, 59(1), 65−86.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

V. ISORELIABILITY MODELS FOR HUMAN SYSTEMS INTEGRATION DOMAIN TRADEOFFS – CHOOSING A PERSONNEL SUPPLY SOURCE FOR FUTURE UNMANNED AIRCRAFT SYSTEM OPERATORS

> Since man is an integral part of the total system, his contributions must be included each and every time that such areas as system performance, system effectiveness, system dependability, system reliability, system capability, and cost effectiveness are considered (Weisz, 1967, p. 3).

A. INTRODUCTION

1. Statement of the Problem

Despite the recognized importance of human systems integration (HSI) domain tradeoffs in enhancing total system performance (Department of Defense [DoD], 1991), current HSI manuals and handbooks do not provide much guidance for making HSI tradeoffs (Barnes & Beevis, 2003). Nor is there a well-established body of knowledge addressing HSI domain tradeoffs despite the obvious need for such information (Barnes & Beevis, 2003). For example, Booher (1990, pp. 12, 42) makes only two references to "tradeoff," and the National Research Council, through its Committee of Human Factors' HSI report (Pew & Mavor, 2007, pp. 3, 19, 34, 140), makes only four references to "tradeoff," none of which even begin to scratch the surface of the issue.

While HSI domains may have important interactions with each other, these interactions are hard to predict and there is little quantitative information to support tradeoff decisions between domains (Barnes & Beevis, 2003). For example, Beevis (1996) analyzed preexisting data on human factors affecting safety and performance in the Canadian F/A-18 Hornet aircraft in an attempt to *qualitatively* assess HSI domain interactions and their impact on performance. Human factors were categorized into HSI domains, and using statistical structural analysis, factor interactions were assessed to identify important direct and indirect domain interactions. The HSI domains were observed to share many statistically significant indirect interactions and relatively few direct interactions, with the personnel domain having the greatest number of interactions

in comparison to the other HSI domains. Overall, the analysis showed that the pattern of HSI domain interactions is complex and does not lead to simple generalizations about tradeoffs among the HSI domains.

In contrast, Jones and Kennedy, along with various other colleagues (Jones, Kennedy, Turnage, Kuntz, & Jones, 1985; Jones, Kennedy, Kuntz, & Baltzley, 1987; Kennedy, Jones, & Baltzley, 1988, 1989; Kennedy & Jones, 1992; Jones, Turnage, & Kennedy, 1993; Jones & Kennedy, 1996; Jones, 2000) advanced a more quantitative tradeoff methodology based on the idea of developing isoperformance curves among the HSI domains. The central reasoning behind their isoperformance methodology is the idea that, almost always, a specified level of performance can be produced by more than one combination of determinants. Hence, the isoperformance curves trace all combinations of two or more HSI domains that produce a specified level of performance. Jones and Kennedy focused on personnel-training interactions and developed their technique primarily to generate tradeoff functions between personnel abilities and factors such as training time and training system effectiveness. However, they emphasize the generalizability of the isoperformance methodology to training, equipment, and manpower tradeoffs.

Only a relative handful of studies have either established a pattern of HSI domain interactions that relate to system performance (Beevis, 1996) or developed quantitative domain tradeoff functions (Jones and Kennedy and colleagues). In the case of Beevis, the interactions examined provided qualitative information about tradeoffs among the HSI domains, but the systems engineering process requires quantitative information for tradeoff analyses. Ideally, such tradeoff analyses should describe equal-cost or equal-performance options (Barnes & Beevis, 2003). The isoperformance methodology advocated by Jones, Kennedy, and colleagues addresses the equal-performance tradeoff analysis, but their only real world isoperformance curves are mainly limited to applications involving personnel and training domain tradeoffs in terms of paper and pencil test performance (Jones & Kennedy, 1992; Jones, 2000). Additionally, their demonstrations of isoperformance curves for HSI applications are of small scale, which is to say single function performance. They do not address the higher dimensionality of the

problem of complex system design (i.e., the curse of dimensionality (Bellman, 1961)), or the need to consider performance across multiple, independent functions. While the isoperformance methodology has been extended to the analysis and design of complex technical systems such as satellite design (de Weck & Jones, 2006), its applicability for more complex HSI problem sets remains an open question.

Unfortunately, most information in the armed services concerning the human system is not well organized or easily located, which hampers program managers, system engineers, and HSI practitioners in meeting military requirements and advancing efficiency goals. Nor do the armed services routinely archive data that would support the generation of isoperformance curves, necessitating that any attempt to construct such curves must be opportunistic (Jones & Kennedy, 1992). This scarcity of data for developing isoperformance curves is regrettable because it is the sort of evidence that program managers and systems engineers appear to require if they are to support the armed services' organizational goals for HSI within their individual system acquisition programs.

This study attempted to contribute to the base of knowledge for HSI domain tradeoffs by exploring the application of the isoperformance methodology to the personnel and training domains in the setting of a multi-dimensional problem. Our study used the opportunistic dataset from the work by Schreiber, Lyon, Martin, and Confer (2002) evaluating the impact of prior flight experience on learning RQ-1 Predator unmanned aircraft system (UAS) operator skills. Schreiber and colleagues' study was conducted to help inform senior decision makers working to develop the best policy for selection and training of Air Force UAS operators. They specifically examined the effect of personnel category, defined in terms of prior flight training and experience, on time to train Predator pilot skills and performance accomplishing a reconnaissance objective. We proposed a simple regression-based analysis for relating the personnel and training domains of HSI to the proportion of proficient people, which allowed us to express human performance probabilistically in terms of functional reliability (Blanchard &

Fabrycky, 2006, pp 369-413). This set the stage for integrating several functional isoreliability models into the systems engineering process using the construct of reliability allocation.

2. Purpose of This Study

The purpose of this study was to adapt the isoperformance concept to relate the HSI personnel and training domains and their interaction to system reliability, using the RQ-1 Predator UAS as our use case. Personnel categories consisted of six groups from which future Air Force UAS operators could potentially be recruited: experienced Air Force pilots, new Air Force fighter/bomber pilots, new Air Force airlift/tanker pilots, civilian instrument-rated private pilots, civilian non-instrument rated private pilots, and Air Force Reserve Officer Training Corps (ROTC) cadets. Participants' training time until reaching criterion performance was examined for three Predator UAS functions: basic maneuver, landing, and reconnaissance. We used data from 93 participants in Schreiber and colleagues' (2002) study to answer these questions:

1) Can we adapt Jones, Kennedy, and colleagues' isoperformance methodology to consider personnel and training domain tradeoffs in terms of the expected proportion of participants that are proficient relative to a fixed level on some performance criterion? That is, can we consider *isoreliability* rather than isoperformance?
2) Can we quantitatively assess the relative importance of the personnel and training domains *and their interaction* in terms of explaining the expected proportion of participants that are proficient for a system function?
3) Can we aggregate our isoreliability curves across system functions to link personnel and training domain considerations to total system reliability?

3. Theoretical Perspective

Conducting personnel and training domain tradeoff analyses—that is, tracing equivalent combinations of personnel qualities and training that yield a specified level of performance—is a highly applied problem. As described by Jones, Kennedy, and Stanney (2004), it is often the case that there are no known empirical regularities or

theories upon which one can rely when making manipulations in the real-world. Accordingly, in practice, system developers and decision makers must hypothesize about human performance-related tradeoffs and then carry out experiments or tests to assess the veracity of their hypotheses. This is not to say that empirical regularities and sound theory are not helpful in performing tradeoff analyses. For example, as was discussed in some detail in Chapter IV, both the power law of practice (i.e., an empirical regularity) and aptitude-treatment interaction (ATI) theory provide us potentially useful insights into the personnel and training factors (determinants) thought to contribute to task performance. However, if control of system performance is to be achieved, it can only be by studying empirically how reliability, or any other "performance" of interest, varies as a function of its determinants in the system-of-interest.

In applying the power law of practice and ATI theory to this study of the acquisition of Predator UAS operator skills, the two major factors in skill acquisition that were recorded by Schreiber and colleagues (2002), practice and experience, are defined in the following manner:

- Practice is the number of trials or total time required to meet criterion performance on a target task.
- Experience is the state of having gained knowledge and skill through direct participation in specific activities related to membership in recognized personnel categories from which future Air Force UAS operators could potentially be recruited.

With these specific definitions, we can now operationalize the HSI training domain in terms of the variable, "practice," and the personnel domain in terms of the variable, "experience." The decision to describe experience in terms of personnel categories relates to the original purpose of Schreiber and colleagues' (2002) study, which was to inform senior decision makers in their selection of a personnel source for the Predator UAS training pipeline. For example, experienced Air Force pilots represent the historical source personnel category. Recent decisions to also use graduates from the two major tracks in the Air Force's specialized undergraduate pilot training pipeline correspond to choices to select from the new Air Force fighter/bomber and airlift/tanker pilot personnel

categories. Other personnel categories considered from a policy perspective have included officer accessions with civilian pilot training and non-pilot officer accessions. Clearly, all these groups differ in multiple dimensions to include human aptitudes, skills, and experiences, but these individual dimensions are too fine grained to be useful variables for senior Air Force leaders on which to base decisions.

The following statement represents the underlying logic for the study. If we specify an *a priori* performance criterion for a task and we can: 1) measure the amount of practice a participant requires to achieve that criterion performance and 2) specify the proportion proficient for a task as a function of practice, experience, and their interaction, then we can develop a quantitative tradeoff function for the HSI training and personnel domains in terms of isoreliability. It is worth calling attention to the fact that isoreliability is not exactly synonymous with Jones, Kennedy, and colleagues' concept of isoperformance, although it borrows heavily from their methodology. Additionally, this study uses a nominal categorical determinant, while Jones, Kennedy, and colleagues only considered integer and continuous determinants. For these reasons, we believe it is worthwhile to discuss here our modification of their isoperformance methodology even though, using their own words, "[isoperformance] does not lead to theory, latent factors, or hypothetical constructs" (Jones & Kennedy, 1996, p. 180).

As previously described, isoperformance is an operational method based on Simon's (1996) notion of "satisficing," and so fixes the amount of performance at an acceptable level and trades off the determinants with respect to each other (de Weck & Jones, 2006). By implication then, while the isoperformance methodology is itself atheoretic, it does presuppose a theoretic causal model between the determinants and the desired performance. The first step in the isoperformance technique is some data-analytic procedure based on a model, ANOVA or regression analyses included. Such a model states the dependent variable(s) as a function of the determinants, parameters to be estimated, and error variations. Once the parameters are estimated, usually using least squares or maximum likelihood statistical techniques, the isoperformance technique requires that the user specify a *criterion* and a level of confidence, which is called the *assurance level*. The criterion is the level of performance desired by the user, and the

assurance level is the probability of attaining that level of performance. Based on the user's choice of criterion and assurance level, the dependent variable is fixed and the resulting equation solved in terms of just the determinants; the determinants can now vary only in ways that will result in the same performance level. In the simple case of just two determinants, plots of every pair of values, each of which will produce the same level of performance, yields an isoperformance curve. Secondary criteria such as cost, safety, or feasibility are then used to identify a preferred solution(s) on the isoperformance curve (Jones & Kennedy, 1992; Jones & Kennedy, 1996; Jones 2000).

In explaining their methodology, Jones, Kennedy, and colleagues often present an example of an isoperformance analysis of the tradeoff between aptitude, as assessed by measured ability, and time on the job for a fixed level of soldier performance, defined in terms of skill qualification test (SQT) scores (Jones & Kennedy, 1996; Jones, 2000). The question they seek to answer is what combinations of aptitude and time on the job are sufficient to achieve a passing SQT score of 60 with 90% certainty. To answer this question, they take the first step of proposing the following model:

$$SQT_{it} = m + T_t + b\left(APT_i - \overline{APT}\right) + cT_t\left(APT_i - \overline{APT}\right) + \varepsilon_{it} \quad (1)$$

where SQT_{it} is the performance of the i^{th} soldier after t years at the entry skill level. APT_i is the soldier's measured ability, and T_t is the effect of time on the job on mean level of performance. $\overline{APT}$ is the mean aptitude score for all soldiers in the dataset and ε_{it} is the normally distributed error term, with mean of zero and variance equal to σ_ε^2. This model allows for two main effects on SQT, aptitude and time on the job, and an interaction between aptitude and time. The model parameters, *m* corresponding to the general mean, *b* the regression coefficient for SQT on aptitude, and *c* the coefficient for the interaction term, along with the error variance, σ_ε^2, are all estimated in the course of fitting the model to the data. The next step is to obtain an expression for the expected performance of the i^{th} soldier:

$$E\left[SQT_{it}\right] = m + T_t + b\left(APT_i - \overline{APT}\right) + cT_t\left(APT_i - \overline{APT}\right) \quad (2)$$

The right hand side of this equation differs from that of the full model only by the absence of the error term, so now the expected performance for the i^{th} soldier depends only on the determinants, APT_i and T_t.

The final step is to determine what the expected performance for the i^{th} soldier must be if the probability of achieving the specified performance is to equal the desired assurance level, in this case 0.90. The performance specifications are met if the expected performance for the i^{th} soldier is:

$$\mathrm{E}\left[SQT_{it}\right] = SQT_{\text{spec}} + z\sigma_{\varepsilon} \tag{3}$$

where SPT_{spec} is the specified level of performance and z equals 1.28 from tables of the standard normal curve.[20] Hence, if SQT score is to equal or exceed 60 with a probability of 0.90, then the expected SQT score for the i^{th} soldier must equal $60 + 1.28\sigma_{\varepsilon}$. The last step is to combine Equations 2 and 3; rearranging terms so that T_t appears as a functions of APT_i:

$$T_t = \frac{SQT_{\text{spec}} + z\sigma_{\varepsilon} - m - b\left(APT_i - \overline{APT}\right)}{1 + c\left(APT_i - \overline{APT}\right)} \tag{4}$$

Using Equation 4 to plot values of the two determinants, aptitude and time on the job, every pair of which produces the same level of performance, results in the isoperformance curve given in Figure V-1.

[20] It is important to note that the assurance level described by Kennedy, Jones, and colleagues ignores an important source of uncertainty, namely that associated with the estimation of the fitted model parameters. Ideally, one should use the prediction interval rather than the confidence interval when establishing the assurance level, because the predictor interval for any setting of the determinants will always be wider than the confidence interval (Montgomery, Peck, & Vining, 2006). However, computation of the prediction interval necessarily requires access to the original data from which the fitted model was derived. For the purpose of historical accuracy, the original description of the method is presented here—and this may well be all that can be done when working from a model published in the literature. However, in subsequent derivations of their method, we make use of the prediction interval unless noted otherwise.

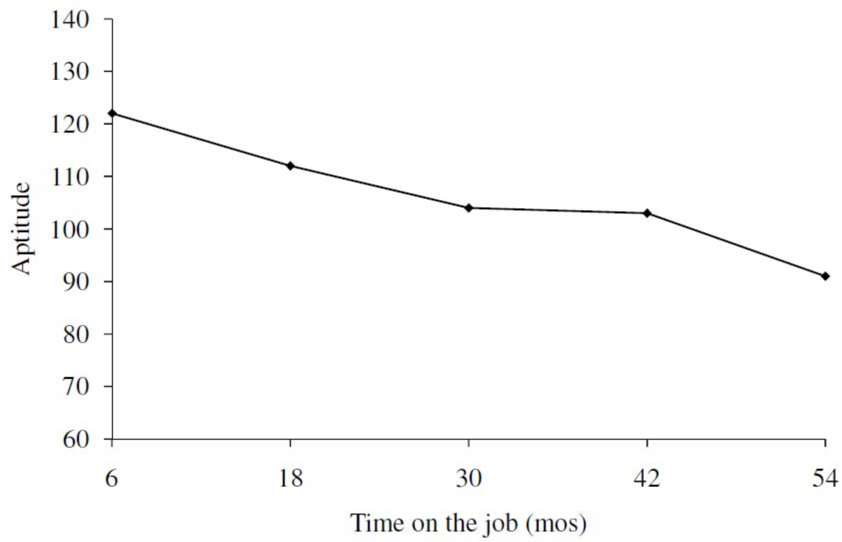


Figure V-1. Isoperformance curve trading off aptitude and time on the job, setting the SQT criterion level at 60 and the assurance level at 0.90 [From Jones & Kennedy, 1996].

In the application of the isoperformance methodology to this study, rather than fixing the level of performance as done by Jones, Kennedy, and colleagues, we propose fixing the proportion of the population of interest that is proficient relative to a reference performance criterion. The question we then seek to answer is what combinations of training and experience are sufficient to achieve a specified proportion proficient with a desired degree of confidence? For the sake of illustration, consider the simple scenario where we have two potential personnel categories, category A and category B, from which we might draw trainees for the job of a system operator. After an arbitrary amount of practice, we can assess whether the i^{th} trainee is proficient. Our response variable, y_i, will take on only two possible values, 1 or 0, depending on whether the i^{th} trainee is or is not proficient respectively. A reasonable probability model for y_i is the binomial with $P\left(y_i = 1\right) = \pi_i$, so to answer our question we start with the following model:

$$y_i = \frac{\exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i}\right)}{1 + \exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i}\right)} + \varepsilon_i \tag{5}$$

where x_{1i} is the number of practice trials accomplished by the i^{th} trainee, x_{2i} is an indicator variable for the i^{th} trainee's personnel category, which takes on a value of one

when the personnel category is B and zero otherwise, and ε_i is the shifted binomially distributed error term with a mean of zero and variance $\sigma_{y_i}^2 = \pi_i\left(1-\pi_i\right)$. This model allows for two main effects on y_i, practice and experience, and an interaction between practice and experience. The βs, or regression coefficients, are parameters estimated in the course of fitting the model to the data. The next step is to obtain an expression for the expected response for the i^{th} trainee:

$$E\left(y_i\right) = \pi_i = \frac{\exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i}\right)}{1+\exp\left(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i}\right)} \tag{6}$$

One should note that the expected response is just the probability that the response variable takes on the value one, which is also the probability the i^{th} trainee is proficient.

Since the probability the i^{th} trainee is proficient is an expected value, which implies an underlying distribution, we need to determine what π_i should be so that the i^{th} trainee achieves a specified probability with a desired assurance level, α:

$$E\left(y_i\right) = \pi_i = \frac{\exp\left(\log\left(\dfrac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + z_\alpha\sqrt{\operatorname{Var}\left(\mathbf{x}_i'\hat{\boldsymbol{\beta}}\right)}\right)}{1+\exp\left(\log\left(\dfrac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + z_\alpha\sqrt{\operatorname{Var}\left(\mathbf{x}_i'\hat{\boldsymbol{\beta}}\right)}\right)} \tag{7}$$

where π_{spec} is the specified probability the i^{th} trainee is proficient, z_α is a lookup from tables of the standard normal curve, and $\operatorname{Var}\left(\mathbf{x}_i'\hat{\boldsymbol{\beta}}\right)$ is the variance of the linear predictor with the predictor expressed in matrix notation (that is $\mathbf{x}_i'\hat{\boldsymbol{\beta}} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_{12} x_1 x_2$). The last step is to combine Equations 6 and 7; rearranging terms:

$$\log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) = \mathbf{x}_i'\hat{\boldsymbol{\beta}} - z_\alpha\sqrt{\operatorname{Var}\left(\mathbf{x}_i'\hat{\boldsymbol{\beta}}\right)} \tag{8}$$

where $\mathbf{x}_i' = \left[1, x_1, x_2, x_1 x_2\right]$ is a vector with $x_1 = \left[0,\infty\right)$ and $x_2 = \left\{0,1\right\}$ and $\hat{\boldsymbol{\beta}}' = \left[\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_{12}\right]$ is a vector of regression coefficients. Observe that the variables

x_1 and x_2 are our two determinants; as will be shown, all the elements in Equation 8 are known, either as constants or as expressions in terms of the determinants.

Everything done up to this point closely mirrors Jones, Kennedy, and colleagues' example analysis of soldier performance. However, we now must address our adaptation to use a logistic regression model with a nominal determinant. First, let us consider the simpler case where a trainee is recruited from personnel category A, in which case $\mathbf{x}_i^{'} = [1, x_1, 0, 0]$. We first calculate the variance of the linear predictor:

$$\mathrm{Var}\left(\mathbf{x}_i^{'}\hat{\boldsymbol{\beta}}\right) = \mathbf{x}_i^{'}\left(\mathbf{X}^{'}\mathbf{V}\mathbf{X}\right)^{-1}\mathbf{x}_i \tag{9}$$

where $\mathbf{X}$ is a matrix of the levels of the regressor variables and V is a diagonal matrix containing the estimated variance of each observation on the main diagonal. The term $\left(\mathbf{X}^{'}\mathbf{V}\mathbf{X}\right)$ is also known as the covariance matrix of the model parameters and can often be obtained directly from commercial statistical software packages. Assuming a given $(\mathbf{X}^{'}\mathbf{V}\mathbf{X})^{-1}$ and computing the variance of the linear predictor in terms of x_1:

$$\begin{aligned}\mathbf{x}_i^{'}\left(\mathbf{X}^{'}\mathbf{V}\mathbf{X}\right)^{-1}\mathbf{x}_i &= [1, x_1, 0, 0]\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix}\begin{bmatrix} 1 \\ x_1 \\ 0 \\ 0 \end{bmatrix} \\ &= a_{1,1} + \left(a_{1,2} + a_{2,1}\right)x_1 + a_{2,2}x_1^2\end{aligned} \tag{10}$$

where the a's are constants representing the values of the elements of the inverted covariance matrix. Substituting this new expression for the variance of the linear predictor into Equation 8 and performing the vector multiplication for $\mathbf{x}_i^{'}\hat{\boldsymbol{\beta}}$:

$$\log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 - z_\alpha\sqrt{a_{1,1} + (a_{1,2} + a_{2,1})x_1 + a_{2,2}x_1^2} \tag{11}$$

Rearranging terms and taking the square of both sides, we can rewrite the above equality as follows:

$$\left(\hat{\beta}_1^2 - z_\alpha^2 a_{2,2}\right) x_1^2 + \left(2\hat{\beta}_0\hat{\beta}_1 - 2\hat{\beta}_1 \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) - z_\alpha^2 a_{1,2} - z_\alpha^2 a_{2,1}\right) x_1$$
$$+\left(\left(\hat{\beta}_0 - \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)\right)^2 - z_\alpha^2 a_{1,1}\right) = 0 \tag{12}$$

Recognizing this is simply a quadratic equation in terms of x_1, we use the quadratic formula and solve for x_1 in terms of the discriminant, Δ:

$$x_1 = f\left(\pi_{\text{spec}}, \alpha \middle| \hat{\beta}_0, \hat{\beta}_1, x_2 = 0\right) = \frac{-2\hat{\beta}_0\hat{\beta}_1 + 2\hat{\beta}_1 \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + z_\alpha^2 a_{1,2} + z_\alpha^2 a_{2,1} \pm \sqrt{\Delta}}{2\left(\hat{\beta}_1^2 - z_\alpha^2 a_{2,2}\right)} \tag{13}$$

where:

$$\Delta = \left(2\hat{\beta}_0\hat{\beta}_1 - 2\hat{\beta}_1 \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) - z_\alpha^2 a_{1,2} - z_\alpha^2 a_{2,1}\right)^2$$
$$-4\left(\hat{\beta}_1^2 - z_\alpha^2 a_{2,2}\right)\left(\left(\hat{\beta}_0 - \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)\right)^2 - z_\alpha^2 a_{1,1}\right) \tag{14}$$

Equation 13 allows us to calculate the minimum necessary training needed for a trainee from personnel category A as a function of both the specified probability proficient and assurance level, given the fitted model parameters.

Now we examine the complementary case where the trainee is recruited from personnel category B, in which case $\mathbf{x}_i' = [1, x_1, 1, x_1]$. Our method remains the same although the calculations become slightly more burdensome. We start again by computing the variance of the linear predictor in terms of x_1, assuming a given $(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}$:

$$
\mathbf{x}_i^{'}\left(\mathbf{X}^{'}\mathbf{V}\mathbf{X}\right)^{-1}\mathbf{x}_i = \left[1, x_1, 1, x_1\right]\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix}\begin{bmatrix} 1 \\ x_1 \\ 1 \\ x_1 \end{bmatrix}
$$

$$
\begin{aligned}
&= \left(a_{1,1} + a_{1,3} + a_{3,1} + a_{3,3}\right) \\
&+ \left(a_{1,2} + a_{1,4} + a_{2,1} + a_{2,3} + a_{3,2} + a_{3,4} + a_{4,1} + a_{4,3}\right)x_1 \\
&+ \left(a_{2,2} + a_{2,4} + a_{4,2} + a_{4,4}\right)x_1^2
\end{aligned} \tag{15}
$$

The result of the matrix multiplication is simply a quadratic in terms of x_1; we propose the following change of variables for simplicity of notation:

$$
\begin{aligned}
c_1 &= a_{1,1} + a_{1,3} + a_{3,1} + a_{3,3} \\
c_2 &= a_{1,2} + a_{1,4} + a_{2,1} + a_{2,3} + a_{3,2} + a_{3,4} + a_{4,1} + a_{4,3} \\
c_3 &= a_{2,2} + a_{2,4} + a_{4,2} + a_{4,4}
\end{aligned} \tag{16}
$$

Substituting our new expression for the variance of the linear predictor into Equation 8, performing the vector multiplication for $\mathbf{x}_i^{'}\hat{\boldsymbol{\beta}}$, and collecting the constant terms:

$$
\log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) = \left(\hat{\beta}_0 + \hat{\beta}_2\right) + \left(\hat{\beta}_1 + \hat{\beta}_{12}\right)x_1 - z_\alpha\sqrt{c_1 + c_2 x_1 + c_3 x_1^2} \tag{17}
$$

By rearranging terms and taking the squares of both sides, we can rewrite the above equality as follows:

$$
\begin{aligned}
&\left(\left(\hat{\beta}_1 + \hat{\beta}_{12}\right)^2 - z_\alpha^2 c_3\right)x_1^2 + \left(2\left(\hat{\beta}_0 + \hat{\beta}_2 - \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)\right)\left(\hat{\beta}_1 + \hat{\beta}_{12}\right) - z_\alpha^2 c_2\right)x_1 \\
&+ \left(\left(\hat{\beta}_0 + \hat{\beta}_2 - \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)\right)^2 - z_\alpha^2 c_1\right) = 0
\end{aligned} \tag{18}
$$

Once again we have a familiar quadratic equation in terms of x_1, allowing us to use the quadratic formula to solve for x_1 in terms of the discriminant, Δ:

$$x_1 = f\left(\pi_{\text{spec}}, \alpha \middle| \hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_{12}, x_2 = 1\right) = \frac{z_\alpha^2 c_2 - 2\left(\hat{\beta}_0 + \hat{\beta}_2 - \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)\right)\left(\hat{\beta}_1 + \hat{\beta}_{12}\right) \pm \sqrt{\Delta}}{2\left(\left(\hat{\beta}_1 + \hat{\beta}_{12}\right)^2 - z_\alpha^2 c_3\right)} \tag{19}$$

where:

$$\begin{aligned}\Delta = &\left(2\left(\hat{\beta}_0 + \hat{\beta}_2 - \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)\right)\left(\hat{\beta}_1 + \hat{\beta}_{12}\right) - z_\alpha^2 c_2\right)^2 \\ &-4\left(\left(\hat{\beta}_1 + \hat{\beta}_{12}\right)^2 - z_\alpha^2 c_3\right)\left(\left(\hat{\beta}_0 + \hat{\beta}_2 - \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)\right)^2 - z_\alpha^2 c_1\right)\end{aligned} \tag{20}$$

Equation 19 allows us to calculate the minimum necessary training needed for a trainee from personnel category B as a function of both the specified proportion proficient and assurance level, given the fitted model parameters. Using Equations 13 and 19, we can compute equivalent combinations of the two determinants, training and personnel category, which yield the same probability that a trainee is proficient.

Recalling that our original question asked about the proportion of the population of interest that is proficient and not the probability that a given trainee is proficient, we now show that these two terms are in fact synonymous:

$$\begin{aligned}\text{Proportion proficient} &= \frac{\text{number of proficient trainees}}{\text{total number of trainees}} \\ &= \frac{\text{probability } i^{\text{th}} \text{ trainee is proficient} \times \text{total number of trainees}}{\text{total number of trainees}} \\ &= \frac{\pi_i n}{n} = \pi_i = \text{probability } i^{\text{th}} \text{ trainee is proficient}\end{aligned}$$

Since *proficiency* is the “present ability to perform representative tasks” (Matthews, Davies, Westerman, & Stammers, 2000, p. 242), the proportion of proficient trainees, π_i, equates to the conditional probability that a new system operator, selected at random from a population of recently trained operators, will satisfactorily accomplish a set of prescribed tasks given some combination of training and prior experience. In view of the

fact that *reliability* may be defined simply “as the probability that a system or product will accomplish its designated mission in a satisfactory manner” (Blanchard & Fabrycky, 2006, p. 370), we propose that the proportion of the population of interest that is proficient is a canonical measure of the reliability of performance for those system functions that are allocated to the human. Hence, Equations 13 and 19 allow us to compute combinations of training and personnel category that are equivalent in terms of reliability—that is, isoreliability curves. Figure V-2 displays a pair of hypothetical isoreliability curves that might be generated using Equations 13 and 19 to tradeoff time spent training and personnel category for two distinct and independent tasks performed by a system operator.

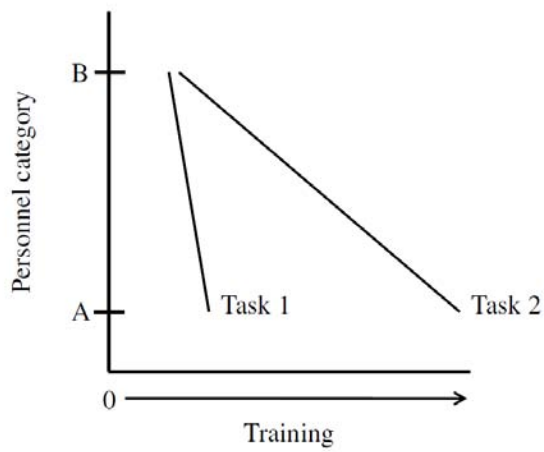


Figure V-2. Hypothetical isoreliability curves trading off time spent training and personnel category for two distinct and independent tasks performed by a system operator.

Figure V-2 illustrates how important information on personnel sensitivity is conveyed using a relatively simple isoreliability readout. Personnel sensitivity, which may include elements of aptitude and prior experience, refers to the general shape of the isoreliability curve. When a curve approaches the horizontal, it is maximally sensitive to personnel category and time spent training is without effect. The opposite extreme, when a curve is vertical, indicates complete insensitivity to personnel category; performance

reliability is determined solely by time spent training. ATI theory suggests that these two extreme cases are unlikely, but they might occur to moderate degrees when decision makers arbitrarily constrain the trade space. For example, a nearly vertical curve, such as is observed in Figure V-2 for task 1, could occur if decision makers considered only closely related personnel categories such as fighter/bomber pilots and tanker/airlift pilots. Likewise, if decision makers sharply truncate the time available for training, it may be that no individuals in certain personnel categories can achieve task proficiency in the allowed range of time spent training. However, the majority of isoreliability curves should resemble that illustrated in Figure V-2 for task 2, where there is a tradeoff between time spent training and personnel category.

While Jones (2000) describes aptitude sensitivity in terms of the first derivate of his isoperformance curves trading off measured ability and training, we cannot quantitatively calculate slopes for our curves because one of our determinants, personnel category, is a categorical variable. Instead, we propose quantitatively considering personnel sensitivity in terms of training time relative to a reference personnel category, much in the same way the indicator variable for personnel category was coded in our logistic regression model:

$$\text{personnel sensitivity} = \frac{T_j - T_i}{T_i} \quad \forall \text{ personnel categories } j \neq i \qquad (21)$$

where T_j is the training time for some personnel category, *j*, and T_i is the training time for a reference personnel category, *i*, such that T_i and T_j are read from the same isoreliability curve. Just as aptitude sensitivity is defined only at a point and may vary from point to point on an isoperformance curve (Jones, 2000), personnel sensitivity is defined for a specific personnel category and may vary across other personnel categories on an isoreliability curve.

While Jones, Kennedy, and colleagues repeatedly discuss the need for tradeoff analysis in addressing many HSI problems, they do not elaborate on how their isoperformance methodology helps systems engineers begin to see human performance and technical tradeoffs together, rather than as a hodge-podge of human factors and

specialty engineering analyses. To help bridge this gap, we begin by looking at Figure V-3, which illustrates the manpower, personnel, and training domains, and by implication of task design, the human factors engineering domain within the context of a system structure (Hay Systems, 1991, as reproduced in Archer, Headley, & Allender, 2003). As a basic system integration model for HSI, Figure V-3 shows that a particular system design concept determines the human tasks that are required, and these tasks in turn drive the requirements for manning, personnel attributes, and needed training. In the description accompanying their model, the original authors suggest several "-ilities," to include reliability, as emergent properties that both result from domain interactions and link individual domain considerations to total system performance:

> Human performance is the product of the interactions of tasks with manpower, personnel, and training. The combination of human performance with the system design, in terms, for example of lethality, mobility, vulnerability, *reliability*, maintainability, and availability drives systems performance [emphasis added] (Hay Systems, 1991, pp. 1,3, as reproduced in Archer, Headley, & Allender, 2003).

The model also implies that the HSI domains, through their contribution to system performance, determine system effectiveness, which is concerned with how well the system performs its mission given the context of the operational environment.

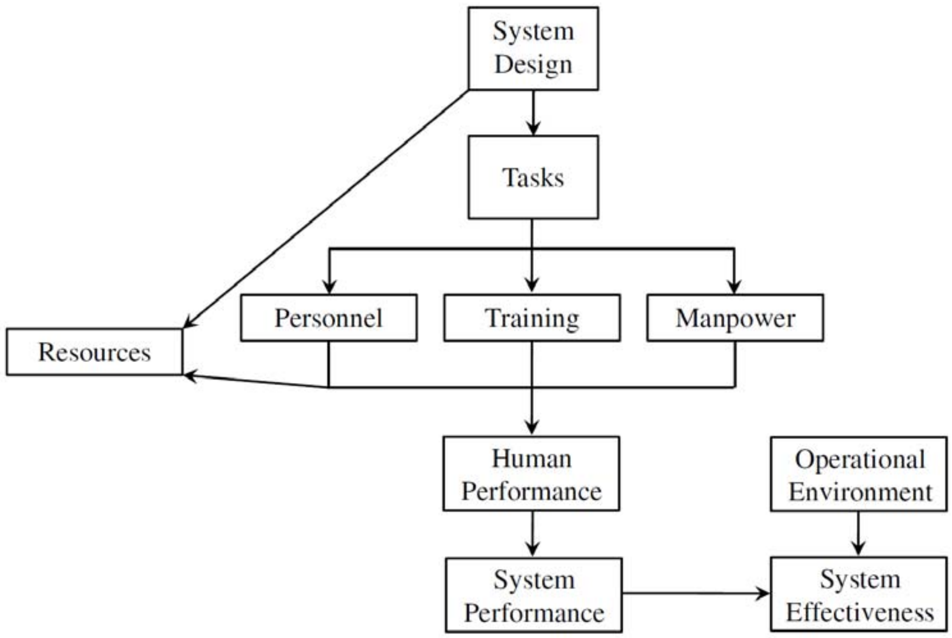


Figure V-3. Manpower, personnel, and training domains within a system structure [From Hay Systems, 1991, as reproduced in Archer, Headley, & Allender, 2003].

In discussing system effectiveness, Pohl (2008) states that "the key system level measures for complex systems that appear in value hierarchies for most complex systems are reliability, availability, and capability" (p. 198). Since capabilities are expressed in terms of system performance thresholds, formulating domain tradeoffs in terms of isoreliability allows us to address two of the three key system level measures while taking advantage of the existing mathematical models in reliability. As mentioned earlier, in simple terms, "reliability is nothing more than the probability that the system under study operates properly for a specified period of time" (Pohl, 2008, p. 198). Mathematical models of reliability focus on items that can be in one of two states at time *t*:

- working $\left(X(t)=1\right)$ and
- not working $\left(X(t)=0\right)$.

Now suppose we have N_0 of such items working at time zero (i.e., $X_i(0)=1,\ i=1$ to N_0) and we define the number of items working at time *t* as $N_s(t)$. If

we let $N_f\left(t\right)$ be a random variable representing the number of items that have failed by time *t*, then $N_f\left(t\right) = N_0 - N_s\left(t\right)$ and the reliability at time *t*, $R\left(t\right)$, can be expressed as:

$$R\left(t\right) = \frac{\mathrm{E}\left[N_s\left(t\right)\right]}{N_0} \tag{22}$$

where $\mathrm{E}\left[N_s\left(t\right)\right]$ is the expected number of items working at time *t*. Likewise, the unreliability of an item, $F\left(t\right)$, can be expressed as:

$$F\left(t\right) = \frac{\mathrm{E}\left[N_f\left(t\right)\right]}{N_0} \tag{23}$$

where $\mathrm{E}\left[N_f\left(t\right)\right]$ is the expected number of items that have failed at time *t*. It should be obvious that we can establish the following relationship:

$$R\left(t\right) = 1 - F\left(t\right) \tag{24}$$

Reliability functions are commonly modeled as continuous time-to-failure distributions. However, certain components or systems have performance characteristics that make it desirable to model their reliability using a discrete distribution. For example, a satellite launch vehicle is better characterized by whether it either launches successfully or does not; time to failure is not an adequate measure to describe the performance of the launch vehicle. Likewise, a pilot's performance is better characterized by whether they successfully land the aircraft or crash. In such cases, an item's (component, subsystem, or system) performance can be characterized in terms of a Bernoulli trial where performance is a random variable that has one of two outcomes: it either works (i.e., a success) or it fails (i.e., a failure) when needed. The probability of success, and hence failure, is constant for each trial, making the binomial or geometric distributions useful for these types of reliability calculations.

We can now consider how formulating domain tradeoffs in terms of reliability allows us to directly link personnel and training domain considerations with total system reliability. Returning to our earlier example of the hypothetical system operator, we focus on the operator as being in one of two states:

- proficient $\left(X\left(x_1, x_2\right)=1\right)$ and
- not proficient $\left(X\left(x_1, x_2\right)=0\right)$

where x_1 is training time and x_2 is personnel category. Now suppose we have $N\left(0, x_2\right)$ initial trainees and we define the number of proficient graduating trainees after some period of training, x_1, as $N_s\left(x_1, x_2\right)$. Consequently, the *human reliability* can be expressed in terms of training time and personnel category as:

$$R_h\left(x_1, x_2\right)=\frac{\mathrm{E}\left[N_s\left(x_1, x_2\right)\right]}{N\left(0, x_2\right)}=\pi_i \tag{25}$$

where π_i is simply the probability the i^{th} trainee is proficient. We can again factor in our assurance level, α, and rewrite Equation 8 to express our human reliability function as follows:

$$R_h\left(x_1, x_2\right)=\frac{1}{1+\exp\left(-\boldsymbol{x}\hat{\boldsymbol{\beta}}+z_\alpha\sqrt{\operatorname{Var}\left(\boldsymbol{x}\hat{\boldsymbol{\beta}}\right)}\right)} \tag{26}$$

It is also possible to define the following relationship:

$$R_h\left(x_1, x_2\right)=1-F_h\left(x_1, x_2\right) \tag{27}$$

where $F_h\left(x_1, x_2\right)$ is the conditional probability that our system operator will fail given a specified training time and personnel category. Hence, $F_h\left(x_1, x_2\right)$ is basically the *human unreliability function*, and it can be defined as follows:

$$F_h\left(x_1, x_2\right)=\frac{\exp\left(-\boldsymbol{x}\hat{\boldsymbol{\beta}}+z_\alpha\sqrt{\operatorname{Var}\left(\boldsymbol{x}\hat{\boldsymbol{\beta}}\right)}\right)}{1+\exp\left(-\boldsymbol{x}\hat{\boldsymbol{\beta}}+z_\alpha\sqrt{\operatorname{Var}\left(\boldsymbol{x}\hat{\boldsymbol{\beta}}\right)}\right)} \tag{28}$$

Since the performance of many system functions involve a human operator interacting with some device, the overall probability that these functions are successfully performed should reflect the contributions of both the human component and the equipment component. Accordingly, the overall reliability (or the probability of successful performance) of a system function, *f*, can be defined as follows:

$$R_f = R_{e,f} \cdot R_{h,f}\left(x_1, x_2\right) \tag{29}$$

where $R_{e,f}$ is the reliability of the equipment used by the system operator in performing function *f* and $R_{h,f}\left(x_1, x_2\right)$ is the human reliability for performing function *f* as previously defined. Figure V-4, taken from Nelson, Schmitz, and Promisel (1984), illustrates how human and equipment reliability affect overall system function reliability. The horizontal and vertical axes represent measures of human and equipment reliability, respectively. Each curve indicates the relationship between human and equipment reliability at certain levels of overall system function reliability. Hence, this figure depicts system function reliability as a mathematical function of both human and equipment reliability.

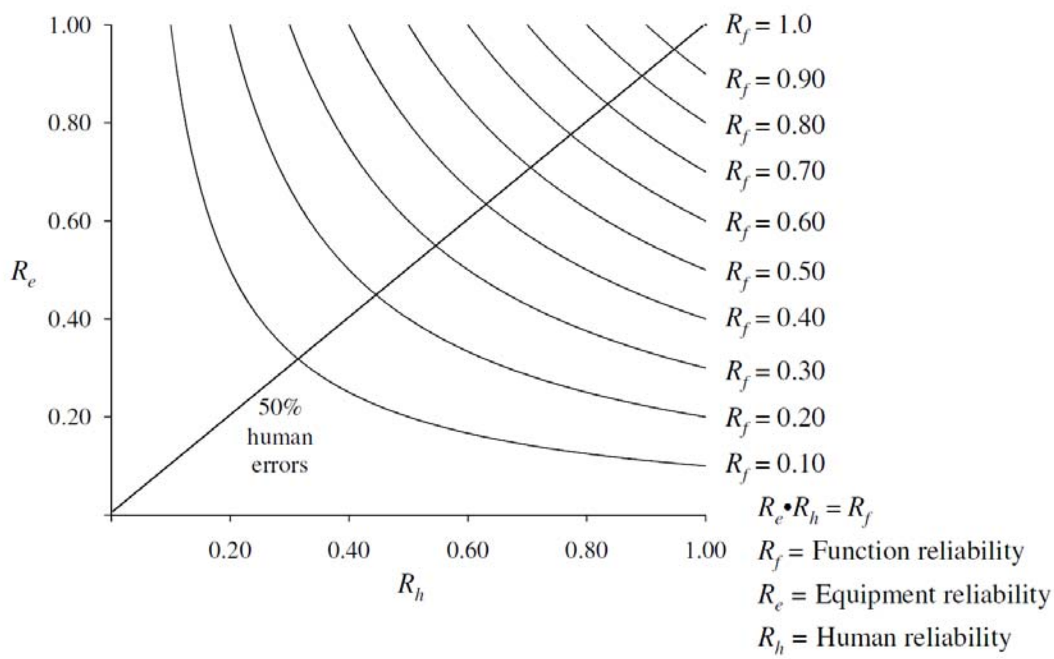


Figure V-4. Effect of human and equipment reliability on overall function reliability [From Nelson, Schmitz, & Promisel, 1984].

A significant advantage in working with reliability rather than performance is the ability to avail ourselves of basic system models. Reliability analysis is generally performed at the lowest levels and results are aggregated into a system level estimate. Usually, a system's functional and physical decompositions can be used to construct a

system level reliability block diagram, the structure of which is used to compute reliability in terms of component and subsystem reliabilities. For example, assume we have a system consisting of *N* functionally independent components, with individual measures of reliability $R_1, R_2, \ldots, R_n$ for some period of time, *t*. If the set of components comprise a series system, then the series system reliability, which is the probability of system success, is given as follows:

$$R_s(t) = R_1(t) \cdot R_2(t) \cdot \ldots \cdot R_n(t) = \prod_{i=1}^{N} R_i(t) \tag{30}$$

If instead, the set of components comprise a parallel system, then the system reliability is given by:

$$R_s(t) = 1 - \left[\left(1 - R_1(t)\right)\left(1 - R_2(t)\right)\ldots\left(1 - R_n(t)\right)\right] = 1 - \prod_{i=1}^{N}\left(1 - R_i(t)\right) \tag{31}$$

K-out-of-*N* systems provide a very general modeling structure in which we assume that a system consists of *N* functionally independent components and the success of the system depends on having at least *K* components operating successfully. Mathematically, this can be modeled as an application of the binomial distribution:

$$R_s = \sum_{i=k}^{N} \binom{N}{i} R^i (1 - R)^{N-i} \tag{32}$$

Overall, most systems are complex combinations of series and parallel system structures of components and subsystems, and system reliability can be constructed by the method of system decomposition.

Although Equation 29 requires reliabilities be specified as probabilities of successful performance, it is entirely possible to express $R_h(x_1, x_2)$ in more familiar terms such as expected (mean) time to failure or failure rate. Remember that $R(x_1, x_2)$ is a probability and it represents the conditional probability that a system operator's performance is satisfactory (i.e., a success) given they have received x_1 training and are from the x_2 personnel category. A similar statement applies to $F(x_1, x_2)$ except that this probability represents the conditional probability that a system operator's performance is unsatisfactory (i.e., a failure). Once we decide on a personnel selection and training

policy, $\left(x_1^*, x_2^*\right)$, we have a fixed probability of success, $p = R\left(x_1^*, x_2^*\right)$, and a fixed probability of failure, $q = F\left(x_1^*, x_2^*\right)$. The geometric distribution is commonly used to model the number of cycles to failure for items that have a fixed probability of failure associated with each cycle. The probability density function for the geometric distribution is given by:

$$P\{N = n\} = (1-q)^{n-1} q \quad \forall\ n = 1, 2, \ldots \tag{33}$$

where N is the number of the cycle on which the first failure occurs. If system cycle lengths, C_i, are independent and identically distributed random variables with an expected cycle length of $\mathrm{E}[C]$, then a reasonable model for the time until the first failure, T, is as follows:

$$T = \sum_{i=1}^{N} C_i \tag{34}$$

The expected time until the first human failure, $\mathrm{E}[T]$, can then be easily computed as follows:

$$\mathrm{E}[T] = \mathrm{E}[N]\mathrm{E}[C] = \frac{1}{q}\mathrm{E}[C] = \frac{1 + \exp\left(-\boldsymbol{x}^*\hat{\boldsymbol{\beta}} + z_\alpha \sqrt{\mathrm{Var}\left(\boldsymbol{x}^*\hat{\boldsymbol{\beta}}\right)}\right)}{\exp\left(-\boldsymbol{x}^*\hat{\boldsymbol{\beta}} + z_\alpha \sqrt{\mathrm{Var}\left(\boldsymbol{x}^*\hat{\boldsymbol{\beta}}\right)}\right)}\mathrm{E}[C] \tag{35}$$

Hence, the expected frequency of system operator failures, $\mathrm{E}[Y]$, or human failure rate, can be expressed as follows:

$$\mathrm{E}[Y] = \frac{1}{\mathrm{E}[T]} = \frac{\exp\left(-\boldsymbol{x}^*\hat{\boldsymbol{\beta}} + z_\alpha \sqrt{\mathrm{Var}\left(\boldsymbol{x}^*\hat{\boldsymbol{\beta}}\right)}\right)}{\left(1 + \exp\left(-\boldsymbol{x}^*\hat{\boldsymbol{\beta}} + z_\alpha \sqrt{\mathrm{Var}\left(\boldsymbol{x}^*\hat{\boldsymbol{\beta}}\right)}\right)\right)\mathrm{E}[C]} \tag{36}$$

We can further extend the concept of human failure rate by next defining a severity rating, s, in terms of the seriousness of the effects or impact of a system operator's failure to satisfactorily perform a function or task. For purpose of illustration, the degree of severity may be expressed quantitatively on a scale of 1 to 10 with regards

to the potential for injury or damage with minor effects being 1, low effects being 2 to 3, moderate effects being 4 to 6, high effects being 7 to 8, and very high effects being 9 to 10 (Blanchard & Fabrycky, 2006, p. 399). We now have the traditional safety and risk management elements of risk likelihood and severity, which can be expressed in terms of a risk assessment value (RAV), defined as:

$$\mathrm{RAV} = \mathrm{E}\left[Y\right] \cdot s \tag{37}$$

where $\mathrm{E}\left[Y\right]$ is the expected failure frequency and s is the failure severity rating. Hence, the RAV for a system function is a possible canonical measure for the safety domain of HSI.

Finally, it is worth taking a moment to note two interesting implications of the preceding discussion for any proposed model of the HSI process. First, the safety domain can be conceptualized as a function of the human factors engineering, personnel, and training domains. Second, safety is probabilistically related to the presence of satisfactory performance, which can be expressed in terms of human reliability, and by way of the compliment, mishaps are probabilistically related to the absence of satisfactory performance. Collectively, these observations suggest a hierarchical relationship of domains as displayed in the updated basic system integration model for HSI depicted in Figure V-5.

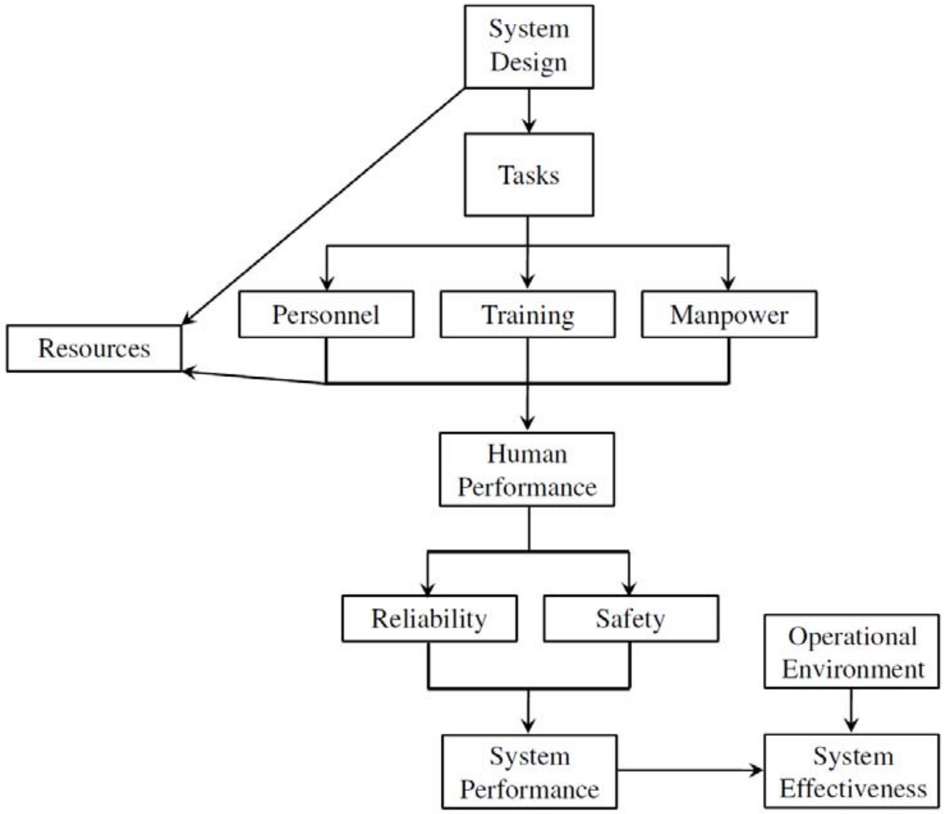


Figure V-5. Manpower, personnel, training, and safety domains within a system structure.

4. Research Questions and Study Hypotheses

The following hypotheses guide the quantitative first phase of this study developing the regression models that are subsequently extended into isoreliability tradeoff functions:

H_1: Personnel category is related to the proportion of participants that are proficient on each of three Predator UAS functions: basic maneuver, landing, and reconnaissance.

H_2: Training time is related to the proportion of participants that are proficient on each of three Predator UAS functions: basic maneuver, landing, and reconnaissance.

H_3: There is an interaction between personnel category and training time in terms of the proportion of participants that are proficient on each of three Predator UAS functions: basic maneuver, landing, and reconnaissance.

The qualitative second phase of this study focuses on the individual isoreliability tradeoff functions and asks two questions:

Q_1: What do the isoreliability tradeoff functions tell us about the relationship between the personnel and training domains of HSI with regards to the RQ-1 Predator UAS mission?

Q_2: Can we aggregate isoreliability tradeoffs across system functions to link personnel and training domain considerations to total system reliability?

5. Definition of Terms

Although we defined terms as they have been introduced in the prior sections, we now provide formal definitions from authoritative sources to strip away any potential multiplicity of meaning from certain words in the interest of precision:

- *Personnel*: Those human aptitudes (i.e., cognitive, physical, and sensory capabilities), knowledge, skills, abilities, and experience levels that are needed to properly perform job tasks (DAU, 2009, p. 7).
- *Reliability*: The ability of a system and its parts to perform its mission without failure, degradation, or demand on the support system (DAU, 2005, B-138).
- *Safety*: Freedom from conditions that can cause death, injury, occupational illness, damage/loss of equipment or property, or damage to the environment (DAU, 2005, p. B-144).
- *Training*: The learning process by which personnel individually or collectively acquire or enhance pre-determined job-relevant knowledge, skills, and abilities by developing their cognitive, physical, sensory, and team dynamic abilities (DAU, 2009, p. 9).

6. Delimitations and Limitations

A delimitation:

We will confine ourselves to the specifics of Schreiber and colleagues' study and their data, which are valid only for the Air Force's RQ-1 Predator mission. Their study focuses on basic aptitudes and skills relevant to piloting the Predator aircraft and does not

address other occupationally significant factors such as leadership, communication skills, general aviation knowledge or familiarity with military operations (Schreiber et al., 2002, pp. 1–2).

A limitation:

Our study involves mining an opportune dataset to identify patterns consistent with HSI domain tradeoff functions (Jones & Kennedy, 1992). Since the original data was collected on convenience samples, the discovery of a particular pattern of domain tradeoffs in the dataset does not necessarily mean that pattern is representative of the whole population from which the data were drawn.

A limitation:

Any patterns discovered in this data could be subject to alternative interpretations.

B. METHODS

1. Participants

The participants were 93 pilots or students expressing a desire to become pilots who were representative of groups from which future Air Force UAS operators could potentially be recruited: experienced Air Force pilots ("Predator selectees"), new Air Force fighter/bomber pilots ("T-38 graduates"), new Air Force airlift/tanker pilots ("T-1 graduates"), civilian instrument-rated private pilots ("Civil instrument pilots"), civilian non-instrument rated private pilots ("Civil private pilots"), and Air Force Reserve Officer Training Corps (ROTC) cadets ("Cadets"). The study used a convenience sample of volunteers recruited from pre-existing, "naturally formed" groups:

1) Predator selectees: Eighteen participants were recruited from the population of experienced Air Force pilots assigned to Predator UAS duties and awaiting the start of their training. Eight participants had experience in fighter/attack aircraft; the remainder had tanker, transport, or bomber aircraft experience. Participants consisted of 17 men, ages ranged from 26 to 43 years, and flight experience ranged from 417 to 3,010 hours with all having at least one prior tour of duty in an operational aircraft squadron.

2) T-38 graduates: Fifteen participants were recruited from the population of students graduating from the Specialized Undergraduate Pilot Training (SUPT) fighter/bomber track at an Air Force training site. Participants consisted of 13 men, ages ranged from 23 to 29 years, and flight experience ranged from 195 to 215 hours, with approximately 120 flight hours in the T-38 Talon.
3) T-1 graduates: Sixteen participants were recruited from the population of students graduating from the SUPT tanker/airlift track at an Air Force training site. Participants consisted of 14 men, ages ranged from 23 to 28 years, and flight experience ranged from 195 to 215 hours, with approximately 105 hours in the T-1 Jayhawk.
4) Civil instrument pilots: Fifteen participants were recruited from the population of pilots recently completing training for an instrument rating at a civil flight school. Participants were all men, ages ranged from 20 to 31 years, and flight experience ranged from 120 to 177 hours, typically in the Cessna model 172 Skyhawk and Beechcraft model 76 Duchess aircraft.
5) Civil private pilots: Thirteen participants were recruited from the population of pilots recently completing training for a single-engine private pilot certificate at a civil flight school. Participants were all men, ages ranged from 18 to 25 years, and flight experience ranged from 45 to 80 hours, typically in the Cessna model 172 Skyhawk aircraft.
6) Cadets: Sixteen participants were recruited from a population of Air Force ROTC cadets at a civilian university. Participants were all men, ages ranged from 19 to 22 years, and none had any fight experience although all intended to pursue Air Force pilot training.

Almost all participants held a bachelor’s or higher degree or were enrolled in an academic program leading to a bachelor’s degree. All active duty military participants, comprising study groups 1-3, took part in the study as part of their normal Air Force duties. The remaining participants, who were civilians, were compensated for their time at the rate of $15 per hour (Schreiber et al., 2002, pp. 3-4).

Six participants were also recruited from the population of experienced Predator UAS pilots at an operational UAS squadron. Participants were all men, ages ranged from 29 to 43 years, and flight experience ranged from 1,680 to 2,942 hours, with approximately 80 to 340 hours flying the MQ-1 Predator UAS (Schreiber et al., 2002, p. 3). This personnel category is not a group from which future Air Force UAS operators could potentially be recruited and so was not included in the study as a comparison group. However, data from these participants was used to establish a performance criterion for one of the study tasks.

2. Research Design

This was a mixed methods study. The first-phase quantitative portion of the study used a quasi-experimental, posttest-only with nonequivalent groups design. The independent variables were defined as follows:

- The categorical variable, *personnel category*, was a measured variable consisting of six levels based on the participants' aviation background: Predator selectees, T-38 graduates, T-1 graduates, civil instrument pilots, civil private pilots, and cadets.
- The continuous variable, *training*, was the treatment variable and was expressed in terms of practice trials or total practice time.

The dependent variable, *proficient*, was a dichotomous variable defined in terms of participant performance relative to the performance criterion set for each experimental task. Participants were classified as "not proficient" if their performance on an experimental task was below the criterion; otherwise, they were classified as "proficient." The second-phase qualitative portion of the study used graphical analysis of isoreliability plots and basic reliability block diagrams.

3. Instruments

Participants' performance was assessed using a modified version of the Air Force Research Laboratory's unmanned aerial vehicle synthetic task environment. This synthetic task environment was based on a simulation of the flight dynamics of the RQ-1A Predator UAS, an early, unarmed variant of the current Predator UAS. The core

aerodynamics model of this simulation was used in a multitask trainer employed by the Air Force to train Predator pilots. Built on top of this Predator model were three synthetic tasks: 1) a basic maneuvering task, 2) a landing task, and 3) a reconnaissance task. These tasks were developed by analyzing real Predator mission tasks and then systematically modifying them to produce synthetic tasks. This was accomplished by conducting extensive structured interviews with expert task performers as part of a cognitive task analysis. The goals, cognitive demands, and required resources for major tasks were identified, and those portions of tasks that were high skill or workload drivers were singled out. These portions were then decoupled from the context of the overall mission for construction into synthetic tasks. The result was a series of synthetic tasks that tapped complex Predator-specific cognitive skills beyond basic stick-and-rudder proficiency, such as sophisticated spatial reasoning and temporal prediction. The overall design philosophy and developmental methodology for the synthetic task environment are described in Martin, Lyon, and Schreiber (1998).

The basic maneuvering task was derived from an instrument flight task designed at the University of Illinois to study expertise-related effects of pilots' visual scan patterns (Wickens, Bellenkes, & Kramer, 1995). The task required participants to fly seven distinct maneuvers while trying to minimize root-mean-squared deviation (RMSD) from ideal performance on airspeed, altitude, and heading. Participants were provided a display of the Predator's legacy head-up display flight symbology overlaid on a black background, hence requiring flight by instrument reference only (Figure V-6). Each maneuver began with a 10-second lead-in during which the participant maintained straight and level flight. A timed maneuver, lasting either 60 or 90 seconds, followed requiring the participant to achieve a target aircraft state by making constant rate changes to one or more of the three flight performance parameters. The initial three maneuvers required the participant to change one flight performance parameter while holding the other two constant. Subsequent maneuvers progressively increased in complexity, requiring the participant to make constant rate changes along two and then three axes of flight. Participants flew each segment repeatedly until simultaneously achieving the

RMSD criterion for all three flight performance parameters. Participants completed the overall task by successfully achieving criterion performance on all seven of the maneuver segments (Schreiber et al., 2002, pp. 7, 40–42)

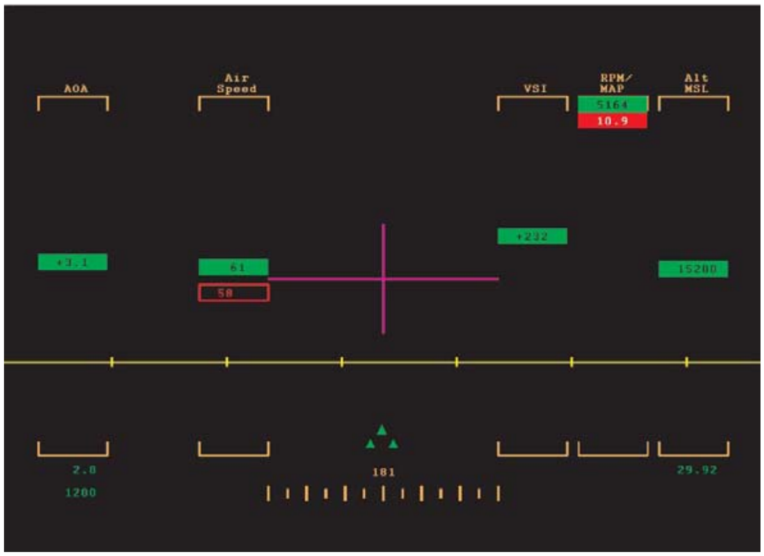


Figure V-6. Display for the synthetic basic maneuvering task [From Schreiber et al., 2002].

The landing task was designed to incorporate many of the challenges of the actual task, such as control latency, high gain of control inputs, aircraft sensitivity to winds, and impoverished sensory feedback in terms of absent vestibular cues and diminished optical flow caused by a limited field-of-view. The task required participants to fly a technical order "typical landing pattern" while trying to meet the criterion on 13 measures of performance: landing pattern ground track RMSD, altitude at three pattern gates, final approach ground track RMSD, final approach glideslope RMSD, touchdown bank angle, touchdown pitch angle, touchdown groundspeed, instantaneous sink rate at touchdown, heading relative to the runway at touchdown, displacement from runway centerline at touchdown, and lateral velocity at touchdown. Each landing trial began with the aircraft located on the downwind leg of the landing pattern and abeam the touchdown point at an altitude of 800 feet above ground level. Participants were provided a display of the Predator's head-up display flight symbology overlaid on simulated imagery from a 30-

degree field-of-view nose camera and a tracker map (Figure V-7). For each trial, participants flew the approach pattern, maintained glideslope, and either landed the aircraft or initiated a go-around. Participants flew repeated landings until simultaneously achieving all 13 criterion measures of performance during a single landing. Participants first achieved criterion performance landing in a no-wind condition; they then achieved criterion performance with a 13-knot crosswind, which was randomly presented during each landing from one of four directions. Participants completed the overall task by successfully achieving criterion performance under both no-wind and crosswind conditions (Schreiber et al., 2002, pp. 7, 43–44).

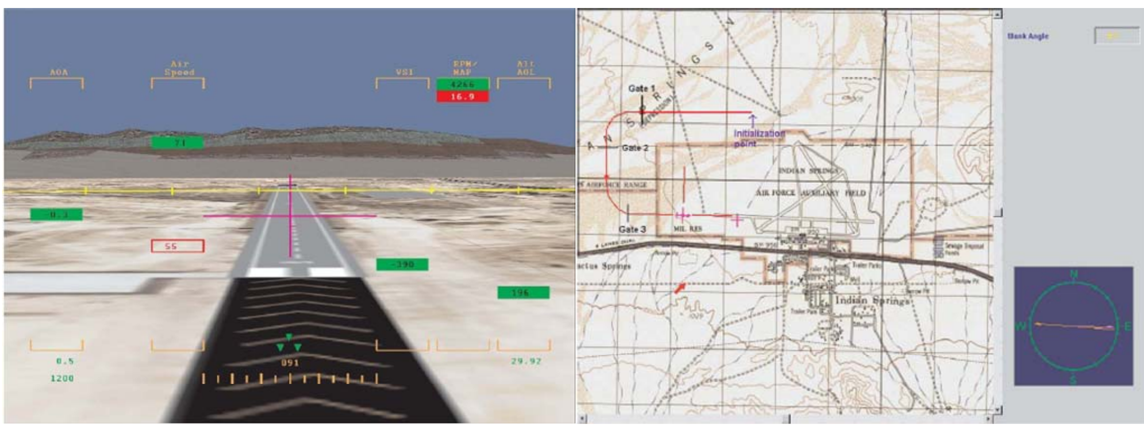


Figure V-7. Primary flight display (left) and tracker map (right) for the synthetic landing task [After Schreiber et al., 2002].

The reconnaissance task was designed to assess participants' aviation-related spatial reasoning skills and was considered the most difficult of the three tasks. The task required participants to maneuver their aircraft so that its payload camera pointed toward a target through a small break in the clouds. Participants were provided a display of the Predator's head-up display flight symbology overlaid on simulated imagery from the payload camera (Figure V-8) or nose camera as well as a tracker map. Participants switched between the nose camera view, to search for the cloud hole, and the payload camera view, to accrue time on the target. Each scenario began with the aircraft on the edge of the target area, which was completely obscured except for a single break in the

clouds. Participants flew 30 scenarios, each with a fixed duration of ten minutes, during which they attempted to maximize their time-on-target while taking into account various constraints such as no-fly zones, altitude restrictions, and camera gimbal limits. Scenarios differed in both wind direction and speed and placement of no-fly zones relative to the cloud hole. The primary measure of performance was a participant's total time-on-target. Violations of constraints resulted in penalty time that was subtracted from total time-on-target (Schreiber et al., 2002, pp. 8–9, 45–47).

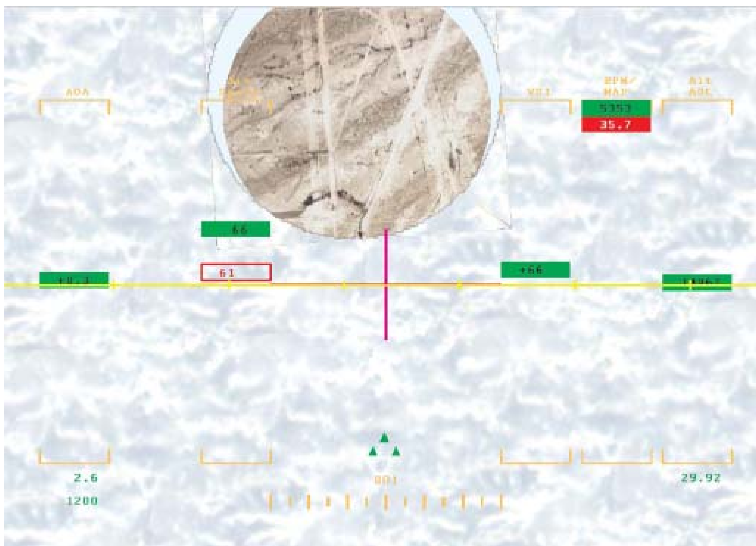


Figure V-8. Primary flight display symbology overlaid on payload camera imagery for the synthetic reconnaissance task [From Schreiber et al., 2002].

The synthetic task environment ran on a dual-Pentium desktop computer networked to a second desktop computer that served as the experimental control station. Figure V-9 illustrates the equipment used in the study. Tasks were presented on two side-by-side 19-inch monitors: the left monitor provided the head-up display flight symbology overlaid on task appropriate background imagery, and the right monitor provided the tracker map display and any additional task relevant information. Participants controlled the simulated Predator aircraft using a joystick, throttle, and rudder pedals. The simulation software was modified to measure participants'

performance on the three synthetic tasks and provided feedback at the end of each trial regarding performance relative to key study parameters (Schreiber et al., 2002, pp. 4–6, 39–40).

Figure V-9. Synthetic task environment equipment setup [From Schreiber et al., 2002].

4. Procedures

All experimental sessions were conducted at the various participant recruiting sites. After being briefed on the study purpose, each participant viewed a self-paced, computer-based tutorial providing declarative and procedural knowledge about the Predator simulation and the particular synthetic task to be flown, starting with the basic maneuvering task. At the end of the tutorial, participants completed a written test; for incorrect responses, participants reviewed the tutorial and reworked the test until obtaining a score of 100 percent. Participants were shown the controls and displays for the basic maneuvering task, provided written reference sheets for the task, and walked through a practice trial by the experimenter with feedback provided using graphical and text displays. Participants then repeated the basic maneuvering task with computerized feedback until criterion performance was achieved. The same general procedure was

used for the landing and reconnaissance tasks: computer-based tutorial, written test, walk through and practice trial, and performance with feedback on the task itself. However, unlike the basic maneuvering and landing tasks, the reconnaissance task was administered for a fixed number of trials rather than until criterion performance was achieved. All participants accomplished the tasks in the same order starting with the basic maneuvering task and finishing with the reconnaissance task. Task sequencing was designed to bring all participants up to a common minimum proficiency on stick-and-rudder skills prior to advancing to the next task. Participants required from 12 to 30 hours to complete the study depending on time spent on tutorial materials and number of trials required to achieve criterion performance on the basic maneuvering and landing tasks (Schreiber et al., 2002, pp. 5–7).

5. Data Analysis Procedures

One of the Air Force Research Laboratory study investigators (D.L.) was contacted and agreed to provide the original study dataset. The dataset was received via e-mail as three Statistical Package for the Social Sciences (SPSS) databases, one for each of the study tasks. Schreiber and colleagues (2002) compared groups on the combined total number of training trials required to achieve criterion performance for the basic maneuvering and landing tasks and total time-on-target for the reconnaissance task (Schreiber et al., 2002, pp. 10–12). Hence, in their study the independent variable was personnel category and the dependent variables were training and time-on-target. Given our theoretical perspective, we extracted data for the independent variables, *personnel category* and *training*, and we calculated a new dependent variable, *proficient*. Data for these variables were copied into version 8.0 of the S-Plus (TIBCO Software Inc., Palo Alto, CA) statistical software package.

Using this opportune dataset, we formulated isoreliability models for each of the study tasks. The general procedure is outlined here, saving discussion of task specific details for the results section. Our dependent or response variable, y_j, took on only two possible values, 1 or 0, depending on whether the j^{th} participant was or was not proficient

respectively. A reasonable probability model for y_j was the binomial with $P\left(y_j = 1\right) = \pi_j$, so we proposed the following model:

$$y_j = \frac{\exp\left(\beta_0 + \sum_{i=1}^{6} \beta_i x_{ij} + \sum_{i=2}^{6} \beta_{1i} x_{1j} x_{ij}\right)}{1 + \exp\left(\beta_0 + \sum_{i=1}^{6} \beta_i x_{ij} + \sum_{i=2}^{6} \beta_{1i} x_{1j} x_{ij}\right)} + \varepsilon_j \tag{38}$$

where x_{1j} was the length of training accomplished by the j^{th} participant, $x_{ij|i\geq 2}$ were indicator variables collectively denoting the personnel category of the j^{th} participant, and ε_j was the shifted binomially distributed error term with a mean of zero and variance $\sigma_{y_j}^2 = \pi_j\left(1 - \pi_j\right)$. Fitting the model to the data using S-Plus, we obtained the following expression for the expected response for the j^{th} participant:

$$E\left(y_j\right) = \pi_j = \frac{\exp\left(\hat{\beta}_0 + \sum_{i=1}^{6} \hat{\beta}_i x_{ij} + \sum_{i=2}^{6} \hat{\beta}_{1i} x_{1j} x_{ij}\right)}{1 + \exp\left(\hat{\beta}_0 + \sum_{i=1}^{6} \hat{\beta}_i x_{ij} + \sum_{i=2}^{6} \hat{\beta}_{1i} x_{1j} x_{ij}\right)} \tag{39}$$

Next we determined what π_j should be so that the j^{th} participant achieved a specified probability with a desired assurance level, α:

$$E\left(y_j\right) = \pi_j = \frac{\exp\left(\log\left(\frac{\pi_{\text{spec}}}{1 - \pi_{\text{spec}}}\right) + z_\alpha \sqrt{\operatorname{Var}\left(\hat{\beta}_0 + \sum_{i=1}^{6} \hat{\beta}_i x_{ij} + \sum_{i=2}^{6} \hat{\beta}_{1i} x_{1j} x_{ij}\right)}\right)}{1 + \exp\left(\log\left(\frac{\pi_{\text{spec}}}{1 - \pi_{\text{spec}}}\right) + z_\alpha \sqrt{\operatorname{Var}\left(\hat{\beta}_0 + \sum_{i=1}^{6} \hat{\beta}_i x_{ij} + \sum_{i=2}^{6} \hat{\beta}_{1i} x_{1j} x_{ij}\right)}\right)} \tag{40}$$

where π_{spec} was the specified probability the j^{th} participant was proficient and z_α was a lookup from tables of the standard normal curve. Combining Equations 39 and 40, rearranging terms, and using matrix notation:

$$\log\left(\frac{\pi_{\text{spec}}}{1 - \pi_{\text{spec}}}\right) = \mathbf{x}_j^{'}\hat{\boldsymbol{\beta}} - z_\alpha \sqrt{\operatorname{Var}\left(\mathbf{x}_j^{'}\hat{\boldsymbol{\beta}}\right)} \tag{41}$$

where $\mathbf{x}_j^{'} = \left[1, x_{1j}, x_{2j}, x_{3j}, x_{4j}, x_{5j}, x_{6j}, x_{1j}x_{2j}, x_{1j}x_{3j}, x_{1j}x_{4j}, x_{1j}x_{5j}, x_{1j}x_{6j}\right]$ was a vector with $x_{1j} = [0, +\infty)$ and $x_{ij} = \{0,1\}$ $\forall$ $i = \{2,3,4,5,6\}$, and $\hat{\boldsymbol{\beta}}$ was a vector of estimated regression coefficients.

We used Equation 41 to express x_{1j} as a function of π_{spec}, α, and personnel category, *k*, such that $x_{i=k,j} = 1$ and $x_{i \neq k,j} = 0$ $\forall$ $k = \{2,3,4,5,6\}$. In effect, we were fixing all the values in Equation 41 other than x_{1j} and simply solving for x_{1j}. Taking the generic case where the personnel category was $k \middle| k \geq 2$, the vector $\mathbf{x}_j^{'} = \left[1, x_{1j} \ldots 1 \ldots x_{1j} \ldots\right]$ had the number one occupying the first and *k*+1th positions and the variable x_{1j} occupying the second and *k*+6th positions in the vector, with the remaining positions simply containing zeros. We first calculated the variance of the linear predictor, $\text{Var}\left(\mathbf{x}_j^{'}\hat{\boldsymbol{\beta}}\right) = \mathbf{x}_j^{'}\left(\mathbf{X}^{'}\mathbf{V}\mathbf{X}\right)^{-1}\mathbf{x}_j$, in terms of x_{1j} given the covariance matrix of model parameters, $\mathbf{X}^{'}\mathbf{V}\mathbf{X}$:

$$
\begin{aligned}
\mathbf{x}_j^{'}\left(\mathbf{X}^{'}\mathbf{V}\mathbf{X}\right)^{-1}\mathbf{x}_j &= \left[1, x_{1j}, \ldots 1 \ldots x_{1j} \ldots\right] \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,12} \\ \vdots & \vdots & \ddots & \vdots \\ a_{12,1} & a_{12,2} & \cdots & a_{12,12} \end{bmatrix} \begin{bmatrix} 1 \\ x_{1j} \\ \vdots \\ 1 \\ \vdots \\ x_{1j} \\ \vdots \end{bmatrix} \\
&= \left(a_{1,1} + a_{k+1,1} + a_{1,k+1} + a_{k+1,k+1}\right) \\
&\quad + \left(a_{2,1} + a_{k+6,1} + a_{1,2} + a_{k+1,2} + a_{2,k+1} + a_{k+6,k+1} + a_{1,k+6} + a_{k+1,k+6}\right) x_{1j} \\
&\quad + \left(a_{2,2} + a_{k+6,2} + a_{2,k+6} + a_{k+6,k+6}\right) x_{1j}^2
\end{aligned} \tag{42}
$$

In the special case where *k* = 1, $\mathbf{x}_j^{'} = \left[1, x_{1j}, 0, \ldots, 0\right]$ and:

$$\mathbf{x}_j^{'}\left(\mathbf{X}^{'}\mathbf{V}\mathbf{X}\right)^{-1}\mathbf{x}_j = \left[1, x_{1j}, 0, \ldots, 0\right]\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,12} \\ \vdots & \vdots & \ddots & \vdots \\ a_{12,1} & a_{12,2} & \cdots & a_{12,12} \end{bmatrix}\begin{bmatrix} 1 \\ x_{1j} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{43}$$

$$= a_{1,1} + \left(a_{1,2} + a_{2,1}\right)x_{1j} + a_{2,2}x_{1j}^2$$

In either case, the result of the matrix multiplication was simply a quadratic in terms of x_{1j}. Continuing with the generic case for the moment, we collected the constants and proposed a change of variables:

$$\begin{aligned} c_{1k} &= a_{1,1} + a_{k+1,1} + a_{1,k+1} + a_{k+1,k+1} \\ c_{2k} &= a_{2,1} + a_{k+6,1} + a_{1,2} + a_{k+1,2} + a_{2,k+1} + a_{k+6,k+1} + a_{1,k+6} + a_{k+1,k+6} \\ c_{3k} &= a_{2,2} + a_{k+6,2} + a_{2,k+6} + a_{k+6,k+6} \end{aligned} \tag{44}$$

This simplified the expression for the variance of the linear predictor, which was now defined as:

$$\mathrm{Var}\left(\mathbf{x}_j^{'}\hat{\boldsymbol{\beta}}\right) = \mathbf{x}_j^{'}\left(\mathbf{X}^{'}\mathbf{V}\mathbf{X}\right)^{-1}\mathbf{x}_j = c_{1k} + c_{2k}x_{1j} + c_{3k}x_{1j}^2 \tag{45}$$

Substituting this new expression into Equation 41 and performing the vector multiplication for $\mathbf{x}_j^{'}\hat{\boldsymbol{\beta}}$:

$$\log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_{1j} + \hat{\beta}_k + \hat{\beta}_{k+5}x_{1j} - z_\alpha\sqrt{c_{1k} + c_{2k}x_{1j} + c_{3k}x_{1j}^2} \tag{46}$$

By rearranging terms and taking the square of both sides, we rewrote the above equality as follows:

$$\begin{aligned} &\left\{\left[\hat{\beta}_0 + \hat{\beta}_k - \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)\right]^2 - z_\alpha^2 c_{1k}\right\} \\ &+\left\{2\left[\hat{\beta}_0 + \hat{\beta}_k - \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)\right]\left(\hat{\beta}_1 + \hat{\beta}_{k+5}\right) - z_\alpha^2 c_{2k}\right\}x_{1j} \\ &+\left[\left(\hat{\beta}_1 + \hat{\beta}_{k+5}\right)^2 - z_\alpha^2 c_{3k}\right]x_{1j}^2 = 0 \end{aligned} \tag{47}$$

Recognizing this as simply a quadratic equation in terms of x_{1j}, we resorted to the quadratic formula and solve for x_{1j} using the discriminant, Δ.

As a result, for each personnel category, *k*, on the ordinate, we calculated corresponding training times of the abscissa that were equivalent in terms of specified reliability and assurance levels. For the special case where *k* = 1:

$$x_{1j} = f\left(\pi_{\text{spec}}, \alpha \middle| \hat{\boldsymbol{\beta}}, \mathbf{X}'\mathbf{V}\mathbf{X}, \mathrm{k}=1\right) = \frac{z_\alpha^2\left(a_{1,2}+a_{2,1}\right)-2\hat{\beta}_1\left[\hat{\beta}_0-\log\left(\dfrac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)\right]+\sqrt{\Delta}}{2\left(\hat{\beta}_1^2-z_\alpha^2 a_{2,2}\right)} \tag{48}$$

where

$$\begin{aligned}\Delta = &\left\{2\hat{\beta}_1\left[\hat{\beta}_0-\log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)\right]-z_\alpha^2\left(a_{1,2}+a_{2,1}\right)\right\}^2 \\ &-4\left(\hat{\beta}_1^2-z_\alpha^2 a_{2,2}\right)\left\{\left[\hat{\beta}_0-\log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)\right]^2-z_\alpha^2 a_{1,1}\right\}\end{aligned} \tag{49}$$

And for the generic case where $k \geq 2$:

$$x_{1j} = f\left(\pi_{\text{spec}}, \alpha \middle| \hat{\boldsymbol{\beta}}, \mathbf{X}'\mathbf{V}\mathbf{X}, \mathrm{k}\geq 2\right) = \frac{z_\alpha^2 c_{2k}-2\left[\hat{\beta}_0+\hat{\beta}_k-\log\left(\dfrac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)\right]\left(\hat{\beta}_1+\hat{\beta}_{k+5}\right)+\sqrt{\Delta}}{2\left[\left(\hat{\beta}_1+\hat{\beta}_{k+5}\right)^2-z_\alpha^2 c_{3k}\right]} \tag{50}$$

where:

$$
\begin{aligned}
c_{1k} &= a_{1,1} + a_{k+1,1} + a_{1,k+1} + a_{k+1,k+1} \\
c_{2k} &= a_{2,1} + a_{k+6,1} + a_{1,2} + a_{k+1,2} + a_{2,k+1} + a_{k+6,k+1} + a_{1,k+6} + a_{k+1,k+6} \\
c_{3k} &= a_{2,2} + a_{k+6,2} + a_{2,k+6} + a_{k+6,k+6} \\
\Delta &= \left\{ 2\left[\hat{\beta}_0 + \hat{\beta}_k - \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}} \right) \right]\left(\hat{\beta}_1 + \hat{\beta}_{k+5} \right) - z_\alpha^2 c_{2k} \right\}^2 \\
&\quad - 4\left[\left(\hat{\beta}_1 + \hat{\beta}_{k+5} \right)^2 - z_\alpha^2 c_{3k} \right]\left\{ \left[\hat{\beta}_0 + \hat{\beta}_k - \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}} \right) \right]^2 - z_\alpha^2 c_{1k} \right\}
\end{aligned}
\tag{51}
$$

These expressions may appear cumbersome, but the values for the many constants were easily obtained from the vector of estimated regression coefficients and the covariance matrix of model parameters provided by S-Plus.

C. RESULTS

1. Basic Maneuvering Task

Schreiber and colleagues present data concerning the number of trials required for participant proficiency on the sequence of seven segments comprising the basic maneuver task. In the study, each iteration of a segment is counted as a trial and participants were required to achieve criterion performance on a segment prior to attempting the next segment in the overall sequence. That is, they repeated a segment and accrued trials until they met criterion performance. Since the temporal length of each segment was fixed but not uniform and the relative difficulty of each segment varied, we computed an overall time to reach proficiency on the basic maneuver task:

$$
x_j = \sum_{s=1}^{7} n_{j,s} t_s \quad \forall\, j \in \{1, 2, \ldots, 93\} \tag{52}
$$

where x_j is the total time for participant *j* to reach proficiency, $n_{j,s}$ is the number of trials needed by participant *j* to reach proficiency on segment *s*, and t_s is the temporal length of segment *s*. We assert that x_j is a better measure of merit than number of trials because it accounts for the fact that the more difficult segments were also the longer segments. A graph of our response variable of interest, *y*, the proportion of participants in each

personnel category who achieve proficiency on the basic maneuver task, versus the total time to reach proficiency (*x*) is shown in Figure V-10.

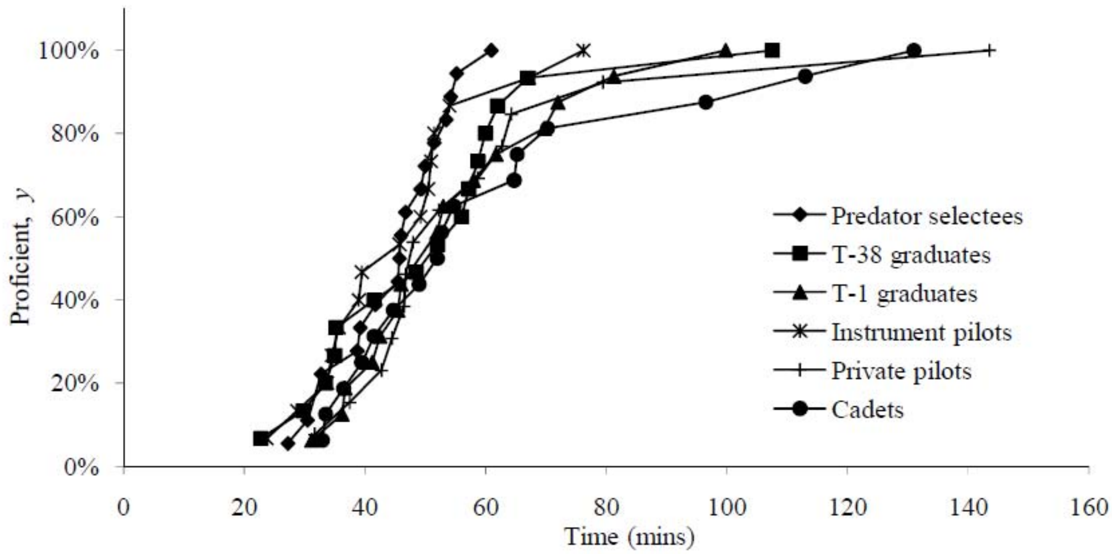


Figure V-10. Scatter plot of the raw basic maneuvering task data.

Given our theoretical perspective, the independent variables were the continuous variable, *Time*, corresponding to the time to reach proficiency and hence the training domain of HSI, and the categorical variable, *Group*, corresponding to personnel category and thus, the personnel domain of HSI. The dependent variable was the proportion proficient, *y*, which was a measure of human performance relative to an *a priori* standard of performance and was the resulting synthesis of the training and personnel domains of HSI. The basic logistic regression model related the proportion proficient to the three potential predictor variables, *Time*, *Group*, and their interaction, $Time \times Group$. The categorical variable, *Group*, was dummy variable coded for inclusion in the regression analysis. Linear and additive regression models, the latter using spline functions to perform piecewise polynomial fitting, were examined and plots of the models were used to assess the fit, determine the influence of outliers, and assure regression assumptions were not violated (see the chapter appendix for details).

Overall, we found that the first order model with interaction based on a linear fit and using the logit link function was the most parsimonious, resulting in the final fitted logistic regression model of:

$$\begin{aligned}\log\left(\frac{\hat{y}}{1-\hat{y}}\right) = & -3.7966 + 0.0708x_1 - 1.6245x_2 - 1.8577x_3 - 3.7123x_4 \\ & -1.4358x_5 - 0.7665x_6 + 0.0555x_1x_2 + 0.0412x_1x_3 + 0.0998x_1x_4 \\ & + 0.0318x_1x_5 + 0.0253x_1x_6\end{aligned} \tag{53}$$

where:

$x_1 =$ Time $[0,+\infty)$

$x_2 =$ Civil instrument pilots $\{0,1\}$

$x_3 =$ Civil private pilots $\{0,1\}$

$x_4 =$ Predator selectees $\{0,1\}$

$x_5 =$ T-1 graduates $\{0,1\}$

$x_6 =$ T-38 graduates $\{0,1\}$

Table V-1 summarizes the estimated regression coefficients and standard errors for the final fitted logistic regression model of the basic maneuvering task data. A graph of the fitted response variable $(\hat{y})$ versus total time to reach proficiency (x) by personnel category is shown in Figure V-11.

Table V-1. Estimated regression coefficients and standard errors for the final fitted model of the basic maneuvering task data.

Variable	$\hat{\beta}$	$\mathrm{se}\left(\hat{\beta}\right)$	Z	p-value
Constant	-3.79656	0.58062	-6.539	<0.0001
Time	0.07078	0.01010	6.440	<0.0001
Group(Civil instrument pilots)	-1.62448	0.98019	-1.657	0.0975
Group(Civil private pilots)	-1.85765	1.18203	-1.572	0.1160
Group((Predator selectees)	-3.71233	1.11643	-3.325	0.0009
Group(T-1 graduates)	-1.43579	0.92641	-1.550	0.1211
Group(T-38 graduates)	-0.76651	0.88977	-0.861	0.3892
Time x Group(Civil instrument pilots)	0.05546	0.02104	2.636	0.0084
Time x Group(Civil private pilots)	0.04121	0.02321	1.776	0.0757
Time x Group(Predator selectees)	0.09976	0.02346	4.252	<0.0001
Time x Group(T-1 graduates)	0.03178	0.01787	1.778	0.0754
Time x Group(T-38 graduates)	0.02533	0.01746	1.451	0.1468

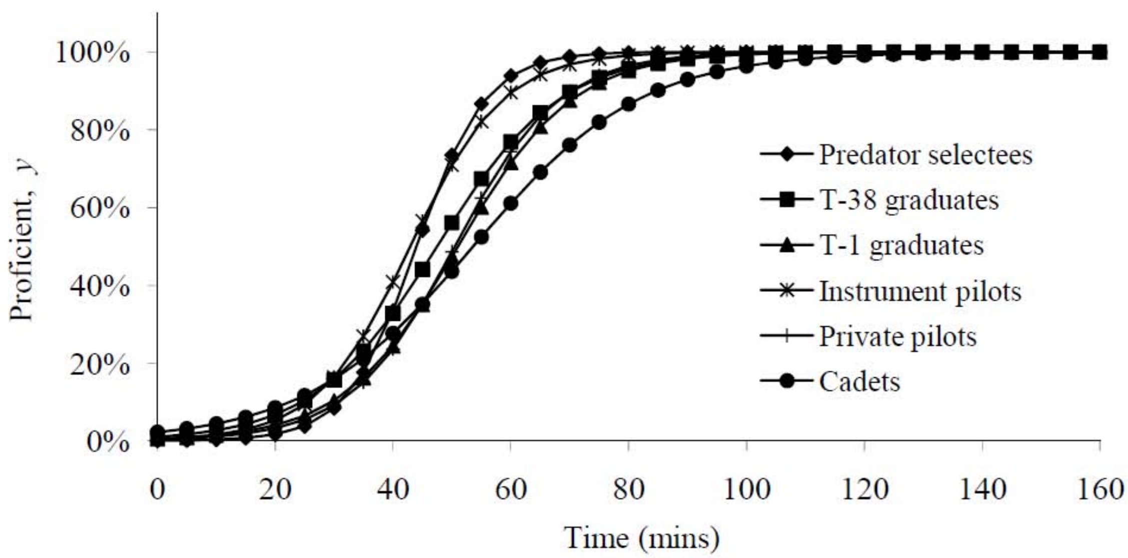


Figure V-11. Plot of the fitted basic maneuvering task data.

We next calculated a system of training-reliability equations, one equation for each personnel category, *k*, from which we then generated the final isoreliability curves. To calculate these training-reliability equations, we used the inverted covariance matrix of model parameters, which was obtained from the S-Plus output:

$$\begin{pmatrix}
0.33713 & -0.00617 & -0.33713 & -0.33713 & -0.33713 & -0.33713 & -0.33713 & 0.00617 & 0.00617 & 0.00617 & 0.00617 & 0.00617 \\
-0.00617 & 0.00012 & 0.00617 & 0.00617 & 0.00617 & 0.00617 & 0.00617 & -0.00012 & -0.00012 & -0.00012 & -0.00012 & -0.00012 \\
-0.33713 & 0.00617 & 0.96078 & 0.33713 & 0.33713 & 0.33713 & 0.33713 & -0.02002 & -0.00617 & -0.00617 & -0.00617 & -0.00617 \\
-0.33713 & 0.00617 & 0.33713 & 1.39720 & 0.33713 & 0.33713 & 0.33713 & -0.00617 & -0.02686 & -0.00617 & -0.00617 & -0.00617 \\
-0.33713 & 0.00617 & 0.33713 & 0.33713 & 1.24643 & 0.33713 & 0.33713 & -0.00617 & -0.00617 & -0.02572 & -0.00617 & -0.00617 \\
-0.33713 & 0.00617 & 0.33713 & 0.33713 & 0.33713 & 0.85824 & 0.33713 & -0.00617 & -0.00617 & -0.00617 & -0.01609 & -0.00617 \\
-0.33713 & 0.00617 & 0.33713 & 0.33713 & 0.33713 & 0.33713 & 0.79169 & -0.00617 & -0.00617 & -0.00617 & -0.00617 & -0.01504 \\
0.00617 & -0.00012 & -0.02002 & -0.00617 & -0.00617 & -0.00617 & -0.00617 & 0.00044 & 0.00012 & 0.00012 & 0.00012 & 0.00012 \\
0.00617 & -0.00012 & -0.00617 & -0.02686 & -0.00617 & -0.00617 & -0.00617 & 0.00012 & 0.00054 & 0.00012 & 0.00012 & 0.00012 \\
0.00617 & -0.00012 & -0.00617 & -0.00617 & -0.02572 & -0.00617 & -0.00617 & 0.00012 & 0.00012 & 0.00055 & 0.00012 & 0.00012 \\
0.00617 & -0.00012 & -0.00617 & -0.00617 & -0.00617 & -0.01609 & -0.00617 & 0.00012 & 0.00012 & 0.00012 & 0.00032 & 0.00012 \\
0.00617 & -0.00012 & -0.00617 & -0.00617 & -0.00617 & -0.00617 & -0.01504 & 0.00012 & 0.00012 & 0.00012 & 0.00012 & 0.00030
\end{pmatrix}$$

Fixing the assurance level at 0.90, we then created the following system of equations:

$$\begin{aligned} x_1 = f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 1\right) &= 53.74591 + 14.71080 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) \\ &+ \sqrt{7.96538 + 3.15762 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 8.57017 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2} \end{aligned} \tag{54}$$

$$\begin{aligned} x_1 = f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 2\right) &= 42.93723 + 8.19299 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) \\ &+ \sqrt{2.83677 - 0.07616 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 2.22566 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2} \end{aligned} \tag{55}$$

$$\begin{aligned} x_1 = f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 3\right) &= 50.54336 + 9.44603 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) \\ &+ \sqrt{4.90116 + 1.03673 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 4.88046 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2} \end{aligned} \tag{56}$$

$$x_1 = f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 4\right) = 43.99287 + 6.00950 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + \sqrt{1.17477 - 0.44239 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 0.87630 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2} \tag{57}$$

$$x_1 = f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 5\right) = 51.05349 + 10.06281 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + \sqrt{4.18379 + 0.69759 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 3.14120 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2} \tag{58}$$

$$x_1 = f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 6\right) = 47.45217 + 10.75663 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + \sqrt{4.90782 - 0.54789 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 3.78492 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2} \tag{59}$$

Solving this system of equations for a particular specified proportion proficient yielded an isoreliability solution set. Figure V-12 provides a graphical display of the resulting isoreliability model for the basic maneuvering task data. The logistic regression analysis of this same data indicates that the beta weights for personnel category and training time were both significant with training time being the more important predictor. The isoreliability curve took the analysis a step further, by tracing all combinations of the two determinants sufficient to provide a specified level of reliability. In this case, these specifications were equivalent to setting the expected proportion proficient equal to 0.50, 0.70, 0.90, 0.95, and 0.99. For example, any combination of personnel category and training time that lay on the curve for 0.95 sufficed to produce an expected proportion proficient of 0.95 with 90% confidence.

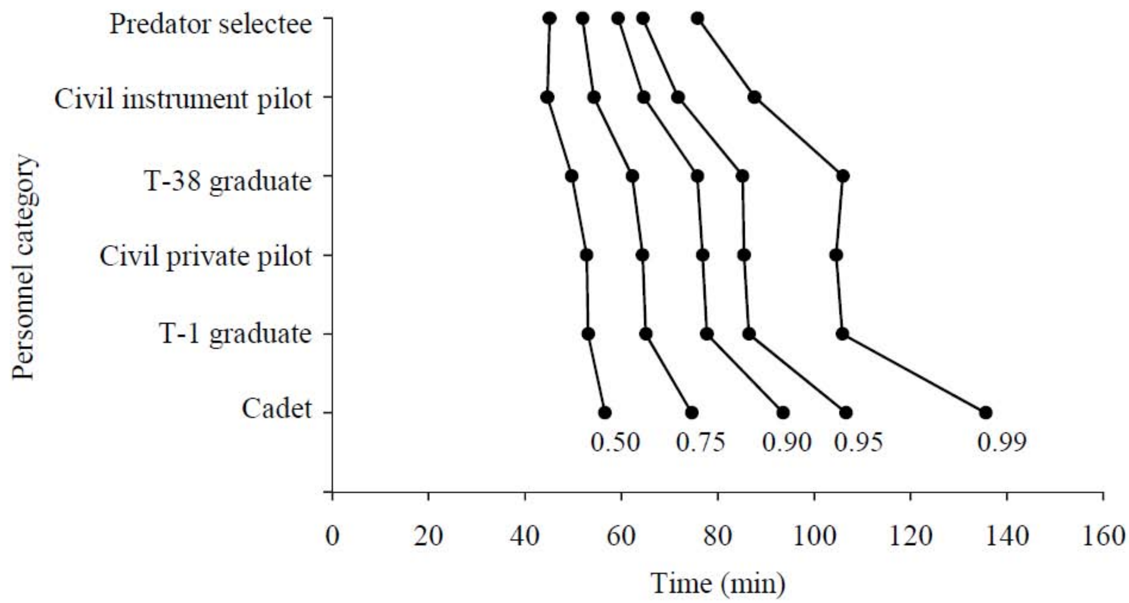


Figure V-12. Isoreliability curves trading off aptitude (personnel category) and training time with the proportion proficient set at 0.50, 0.75, 0.90, 0.95, and 0.99 and level of assurance set at 0.90 for all criterion settings.

Figure V-13 displays the percent difference in training time for each personnel category relative to a reference personnel category, in this case Predator selectees, for various settings of the reliability criterion—a construct that was previously defined as *personnel sensitivity*. Reliability on the basic maneuvering task appeared insensitive to personnel category when the criterion level was set relatively low, that is 0.50. However, increasing the criterion level was associated with a monotonic increase in personnel sensitivity, although there was a divergent pattern across subsets of personnel categories. For instance, personnel sensitivity increased relatively sharply at higher reliability criterion levels for the cadet category but was relatively flat across the range of criterion levels for the civil instrument pilot category. Between these two extremes was the set of personnel categories comprised of civil private pilots, T-38 graduates, and T-1 graduates, which individually appeared insensitive relative to each other across the range of criterion levels.

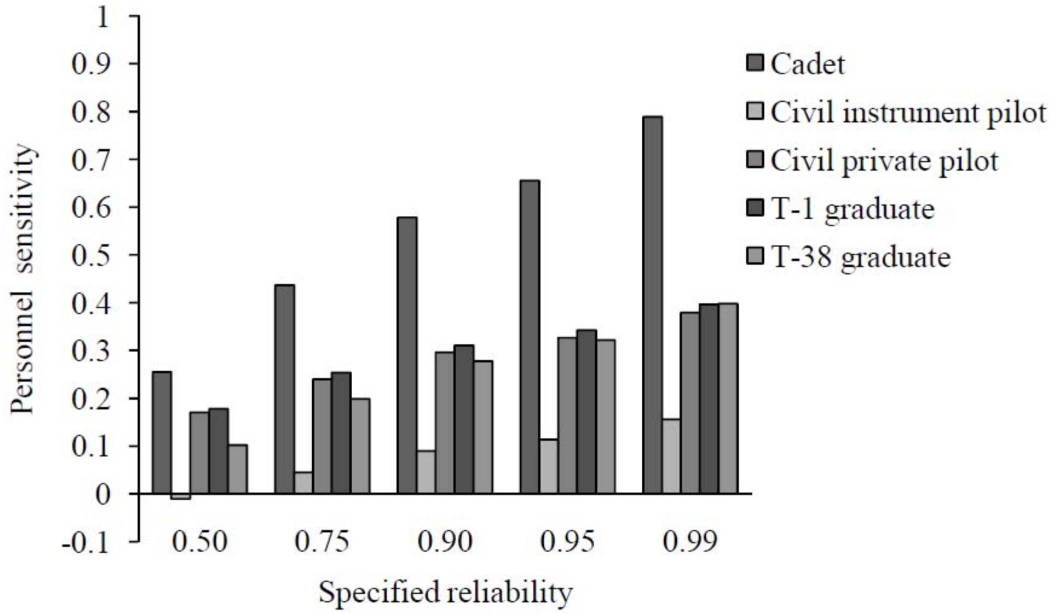


Figure V-13. Personnel sensitivity, expressed in terms of the percent difference in training time for each personnel category relative to the Predator selectee category, for various settings of the reliability criterion on the basic maneuvering task.

2. Landing Task

Schreiber and colleagues presented data concerning 13 criteria assessed during the landing task and the number of trials required until proficiency simultaneously was achieved on all criteria. The response variable of interest, y, was the proportion of participants in each personnel category who achieved proficiency on the landing task. A graph of the response variable versus the number of trials is shown in Figure V-14. As in the case of the basic maneuvering task, a reasonable probability model for the number of proficient participants was the binomial; therefore a logistic regression model was fitted to the data.

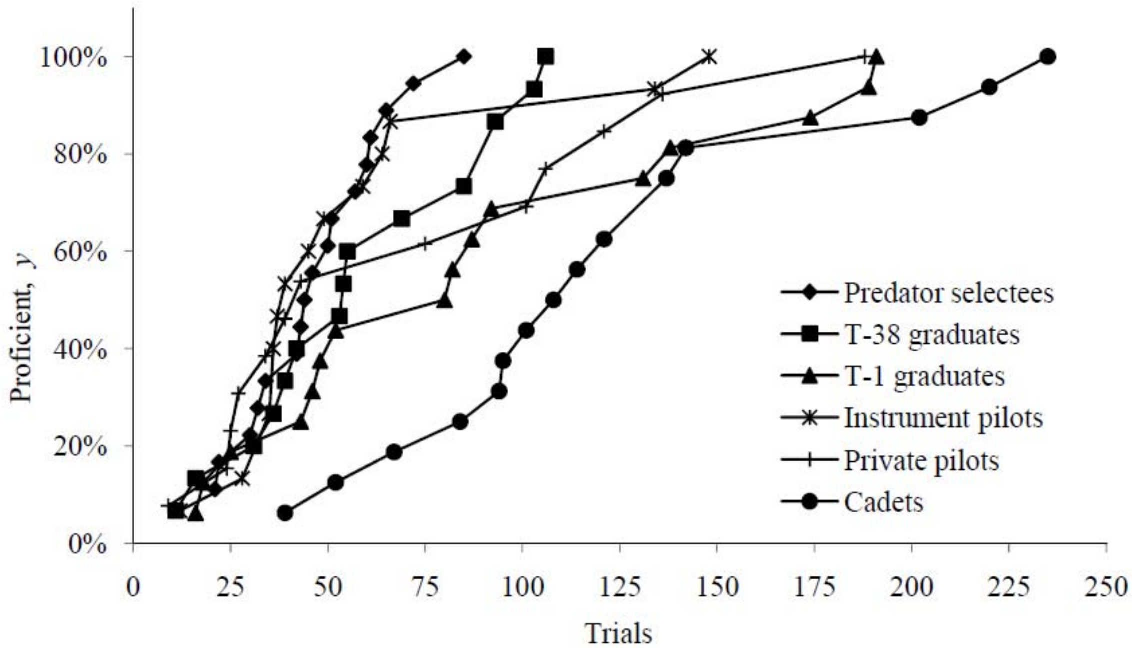


Figure V-14. Scatter plot of the landing task data.

The independent variables were the continuous variable, *Trial*, corresponding to the number of trials to reach proficiency and the categorical variable, *Group*, corresponding to personnel category. The dependent variable was the proportion proficient, *y*, which was a measure of human performance on the landing task relative to an *a priori* standard of performance. The basic logistic model related the proportion proficient to the three potential predictor variables, *Trial*, *Group*, and their interaction, $Trial \times Group$. The categorical variable, *Group*, was dummy variable coded for inclusion in the regression analysis. Linear and additive regression models were examined and plots of the models were used to assess the fit, determine the influence of outliers, and assure regression assumptions were not violated (see the chapter appendix for details).

We again found that the first order model with interaction based on a linear fit and using the logit link function was the most parsimonious, resulting in the final fitted logistic regression model of:

$$\log\left(\frac{\hat{y}}{1-\hat{y}}\right) = -3.5976 + 0.0325x_1 + 0.8351x_2 + 1.8278x_3 - 0.4717x_4 + 1.5178x_5 + 0.7784x_6 + 0.0317x_1x_2 - 0.0025x_1x_3 + 0.0531x_1x_4 - 0.0054x_1x_5 + 0.0200x_1x_6 \qquad (60)$$

where:

$x_1 =$ Trials $[0,+\infty)$

$x_2 =$ Civil instrument pilots $\{0,1\}$

$x_3 =$ Civil private pilots $\{0,1\}$

$x_4 =$ Predator selectees $\{0,1\}$

$x_5 =$ T-1 graduates $\{0,1\}$

$x_6 =$ T-38 graduates $\{0,1\}$

Table V-2 summarizes the estimated regression coefficients and standard errors for the final fitted logistic regression model of the landing task data. A graph of the fitted response variable $(\hat{y})$ versus number of trials to reach proficiency (x) by personnel category is shown in Figure V-15.

Table V-2. Estimated regression coefficients and standard errors for the final fitted model of the landing task data.

Variable	$\hat{\beta}$	$\text{se}\left(\hat{\beta}\right)$	Z	p-value
Constant	-3.59761	0.53662	-6.704	<0.0001
Trial	0.03254	0.00483	6.737	<0.0001
Group(Civil instrument pilots)	0.83514	0.77583	1.076	0.2819
Group(Civil private pilots)	1.82784	0.62964	2.903	0.0037
Group((Predator selectees)	-0.47172	0.74527	-0.633	0.5267
Group(T-1 graduates)	1.51785	0.61675	2.461	0.0139
Group(T-38 graduates)	0.77844	0.68520	1.136	0.2560
Trial x Group(Civil instrument pilots)	0.03169	0.01330	2.383	0.0172
Trial x Group(Civil private pilots)	-0.00251	0.00684	-0.367	0.7136
Trial x Group(Predator selectees)	0.05310	0.01163	4.564	<0.0001
Trial x Group(T-1 graduates)	-0.00542	0.00602	-0.900	0.3681
Trial x Group(T-38 graduates)	0.01998	0.00891	2.242	0.0250

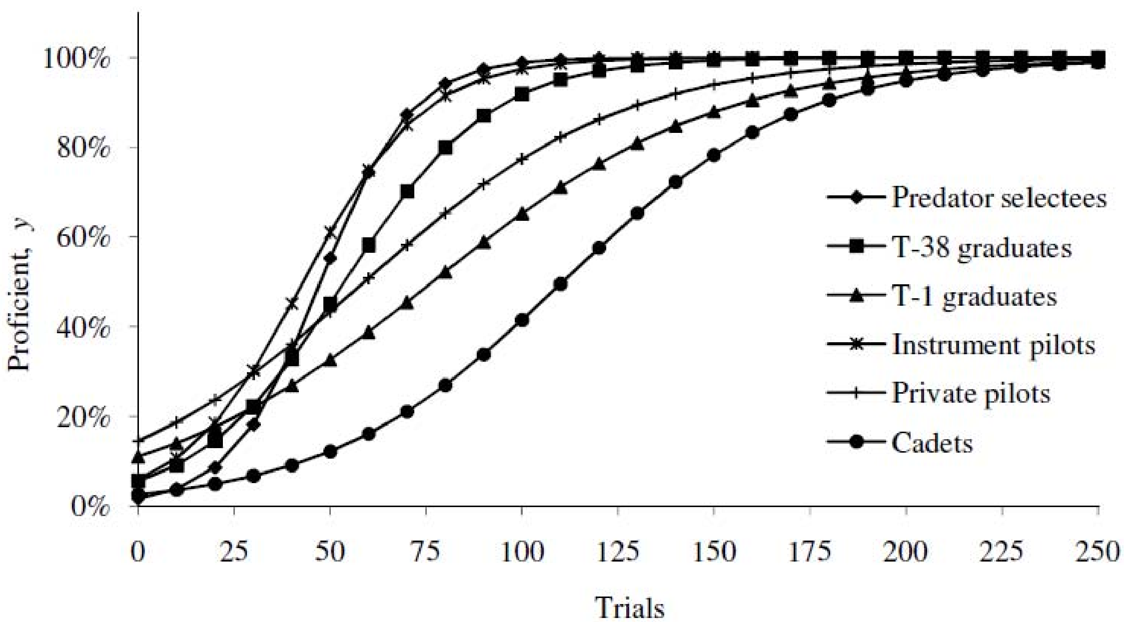


Figure V-15. Plot of the fitted landing task data.

We again calculated a system of training-reliability equations—one equation for each personnel category—using the coefficient vector and the inverted covariance matrix of model parameters, the latter of which is shown below:

$$
\begin{pmatrix}
0.28796 & -0.00248 & -0.28796 & -0.28796 & -0.28796 & -0.28796 & -0.28796 & 0.00248 & 0.00248 & 0.00248 & 0.00248 & 0.00248 \\
-0.00248 & 0.00002 & 0.00248 & 0.00248 & 0.00248 & 0.00248 & 0.00248 & -0.00002 & -0.00002 & -0.00002 & -0.00002 & -0.00002 \\
-0.28796 & 0.00248 & 0.60191 & 0.28796 & 0.28796 & 0.28796 & 0.28796 & -0.00909 & -0.00248 & -0.00248 & -0.00248 & -0.00248 \\
-0.28796 & 0.00248 & 0.28796 & 0.39644 & 0.28796 & 0.28796 & 0.28796 & -0.00248 & -0.00379 & -0.00248 & -0.00248 & -0.00248 \\
-0.28796 & 0.00248 & 0.28796 & 0.28796 & 0.55542 & 0.28796 & 0.28796 & -0.00248 & -0.00248 & -0.00776 & -0.00248 & -0.00248 \\
-0.28796 & 0.00248 & 0.28796 & 0.28796 & 0.28796 & 0.38038 & 0.28796 & -0.00248 & -0.00248 & -0.00248 & -0.00342 & -0.00248 \\
-0.28796 & 0.00248 & 0.28796 & 0.28796 & 0.28796 & 0.28796 & 0.46950 & -0.00248 & -0.00248 & -0.00248 & -0.00248 & -0.00541 \\
0.00248 & -0.00002 & -0.00909 & -0.00248 & -0.00248 & -0.00248 & -0.00248 & 0.00018 & 0.00002 & 0.00002 & 0.00002 & 0.00002 \\
0.00248 & -0.00002 & -0.00248 & -0.00379 & -0.00248 & -0.00248 & -0.00248 & 0.00002 & 0.00005 & 0.00002 & 0.00002 & 0.00002 \\
0.00248 & -0.00002 & -0.00248 & -0.00248 & -0.00776 & -0.00248 & -0.00248 & 0.00002 & 0.00002 & 0.00014 & 0.00002 & 0.00002 \\
0.00248 & -0.00002 & -0.00248 & -0.00248 & -0.00248 & -0.00342 & -0.00248 & 0.00002 & 0.00002 & 0.00002 & 0.00004 & 0.00002 \\
0.00248 & -0.00002 & -0.00248 & -0.00248 & -0.00248 & -0.00248 & -0.00541 & 0.00002 & 0.00002 & 0.00002 & 0.00002 & 0.00008
\end{pmatrix}
$$

Fixing the assurance level at 0.90, we then created the following system of equations:

$$
\begin{aligned}
x_1 = f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 1\right) &= 110.73590 + 31.88856 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) \\
&+ \sqrt{40.89798 + 10.48375 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 36.79571 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2}
\end{aligned} \tag{61}
$$

$$
\begin{aligned}
x_1 = f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 2\right) &= 43.01196 + 16.58483 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) \\
&+ \sqrt{12.46994 - 0.08121 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 16.81486 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2}
\end{aligned} \tag{62}
$$

$$
\begin{aligned}
x_1 = f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 3\right) &= 59.06788 + 34.78748 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) \\
&+ \sqrt{66.83660 + 9.03469 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 51.65751 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2}
\end{aligned} \tag{63}
$$

$$x_1 = f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 4\right) = 47.52709 + 11.97743 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + \sqrt{4.25134 + 0.23100 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 3.59887 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2} \tag{64}$$

$$x_1 = f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 5\right) = 76.79456 + 37.97085 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + \sqrt{53.49178 + 7.77132 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 41.59673 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2} \tag{65}$$

$$x_1 = f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 6\right) = 53.73450 + 19.70057 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + \sqrt{17.63044 + 1.95638 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 12.95919 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2} \tag{66}$$

Using the same method described for the basic maneuver task data, we traced all combinations of the two determinants, training and personnel, sufficient to provide a specified level of performance. We set these specifications, given in terms of proportion proficient, equal to 0.50, 0.70, 0.90, 0.95, and 0.99. Figure V-16 provides a graphical display of the resulting isoreliability model for the landing task data. The logistic regression analysis of this same data indicates that the beta weights for personnel category and number of trials (i.e., training) were both significant with training being the more important predictor. The isoreliability curves showed that the landing task, in contrast to the basic maneuvering task, was more sensitive to personnel factors.

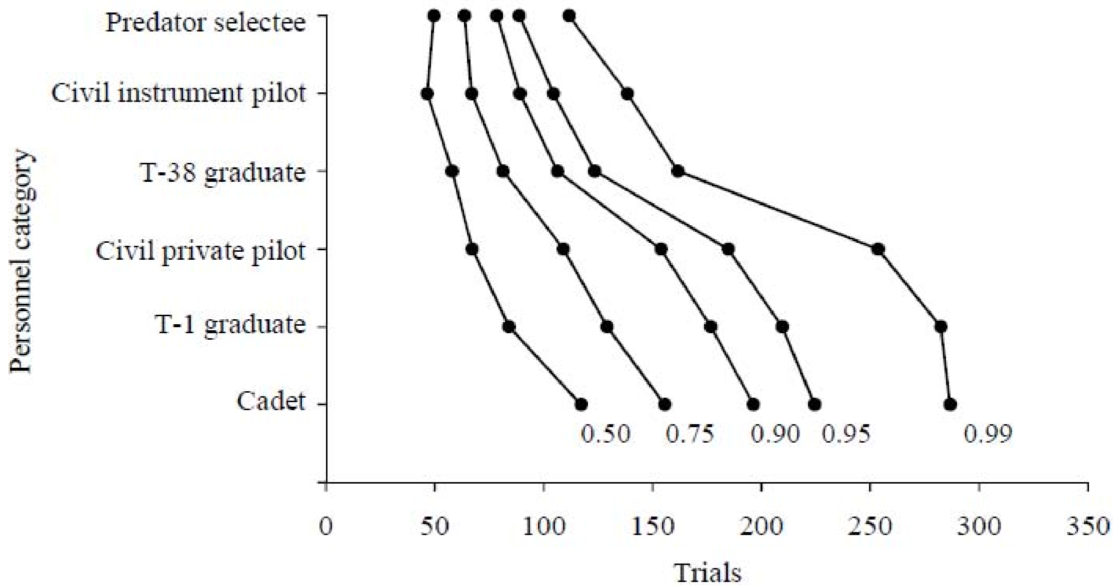


Figure V-16. Isoreliability curves trading off aptitude (personnel category) and training trials with the proportion proficient set at 0.50, 0.75, 0.90, 0.95, and 0.99 and level of assurance set at 0.90 for all criterion settings.

Figure V-17 displays the personnel sensitivity, expressed in terms of the percent difference in training time for each personnel category relative to the Predator selectee category, for various settings of the reliability criterion. Making the criterion for reliability more stringent, which equates to moving rightward across the graphical display, was associated with an increase in the relative personnel sensitivity of the landing task. We saw distinct monotonically increasing trends in personnel sensitivity between the set of categories comprised of civil instrument pilots and T-38 graduates, the set formed of civil private pilots and T-1 graduates, and the set consisting solely of cadets. It is interesting to note that the personnel sensitivity of civil private pilots and T-1 graduates approached that of cadets at higher reliability criterion levels.

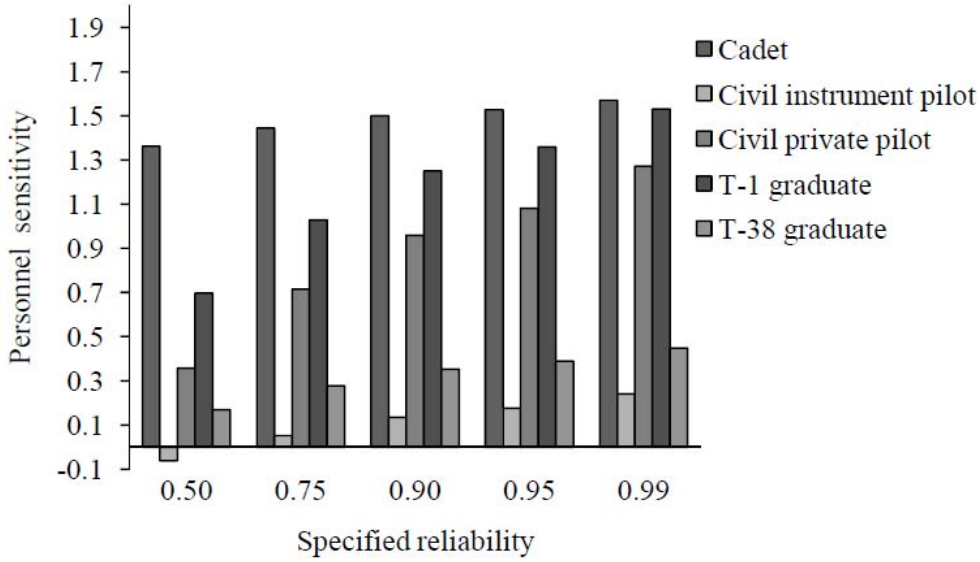


Figure V-17. Personnel sensitivity, expressed in terms of the percent difference in training time for each personnel category relative to the Predator selectee category, for various settings of the reliability criterion on the landing task.

3. Reconnaissance Task

Schreiber and colleagues (2002) presented data on the total time on target for each trial, with all participants completing thirty 10-minute trials. Unlike the previous tasks, no *a priori* performance criterion was specified in the original study, hence requiring us to develop a criterion given the available data. This task was accomplished using the data collected from six experienced Predator pilots. Figure V-18 displays a scatter plot, by pilot, of the total time for each trial that the sensor camera was viewing the target through the cloud hole. We next checked for a bivariate relationship between *Time* and *Trial* using the rank-based Spearman's measure of correlation. Although it was not obvious in the scatter plot, there was a weak positive correlation (ρ = 0.196, p-value = 0.009) between *Time* and *Trial*. Since we expected that the performance of experienced Predator pilots should approach an asymptote over the 30 trials, the data were fit using both a logarithmic and an inverse model.

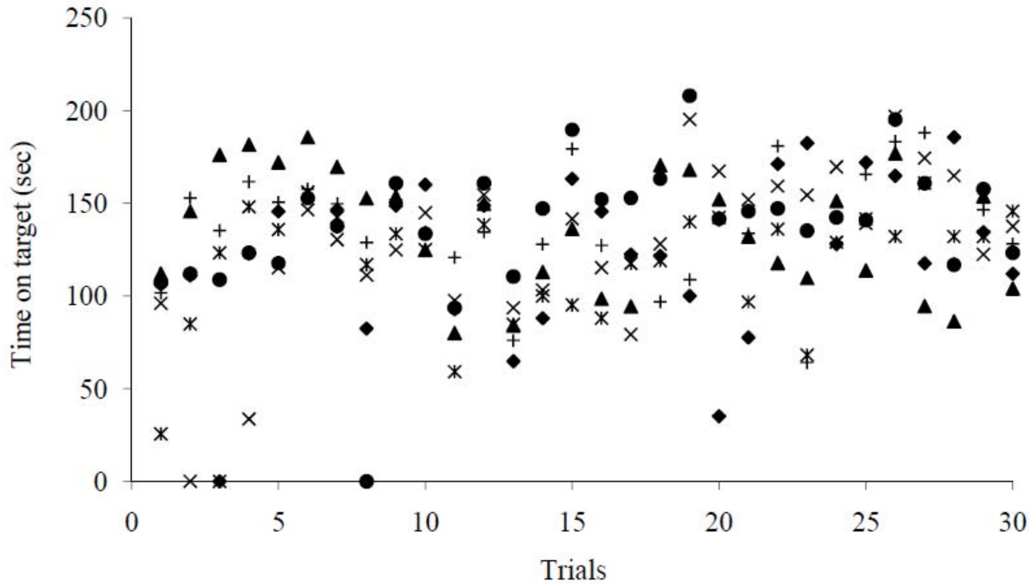


Figure V-18. Scatter plot of the reconnaissance task data for six Predator pilots.

Our initial regression models related *Time* to the predictor variable, *Trial*, using either a logarithmic and inverse transformation of *Trial*. Table V-3 summarizes the fit of these linear models and Figure V-19 displays the predicted observations from the two models relative to a scatter plot of the actual observations. Both models appeared comparable, so we selected the inverse model as it had the advantage over the logarithmic model of a finite limit:

$$\begin{aligned} &Time = f(Trial) = 98.4929 + 12.4195 \times \log(Trial) \\ &\lim_{Trial \to \infty} f(Trial) = \infty \end{aligned} \tag{67}$$

$$\begin{aligned} &Time = f(Trial) = 136.5474 - 53.6727 \times \frac{1}{Trial} \\ &\lim_{Trial \to \infty} f(Trial) = 136.5474 \end{aligned} \tag{68}$$

The asymptotic limit of the average performance of Predator pilots, namely 136.547 seconds on target, was used henceforth as the criterion level of performance for the reconnaissance task.

Table V-3. Estimated regression coefficients and standard errors for the final fitted model of the landing task data.

	Logarithmic model†			Inverse model‡		
Variable	$\hat{\beta}$	$\mathrm{se}(\hat{\beta})$	*t*-value	$\hat{\beta}$	$\mathrm{se}(\hat{\beta})$	*t*-value
Constant	98.4929	8.5417	11.5308*	136.5475	3.3287	41.0207*
log(*Trial*)	12.4195	3.2536	3.8172*	---	---	---
$Trial^{-1}$	---	---	---	-53.6727	14.3593	-3.7378*

†$F_{1,178} = 14.57$, $p = 0.0002$, $R^2 = 0.07566$
‡$F_{1,178} = 13.97$, $p = 0.0003$, $R^2 = 0.07278$
*$p \leq 0.0002$

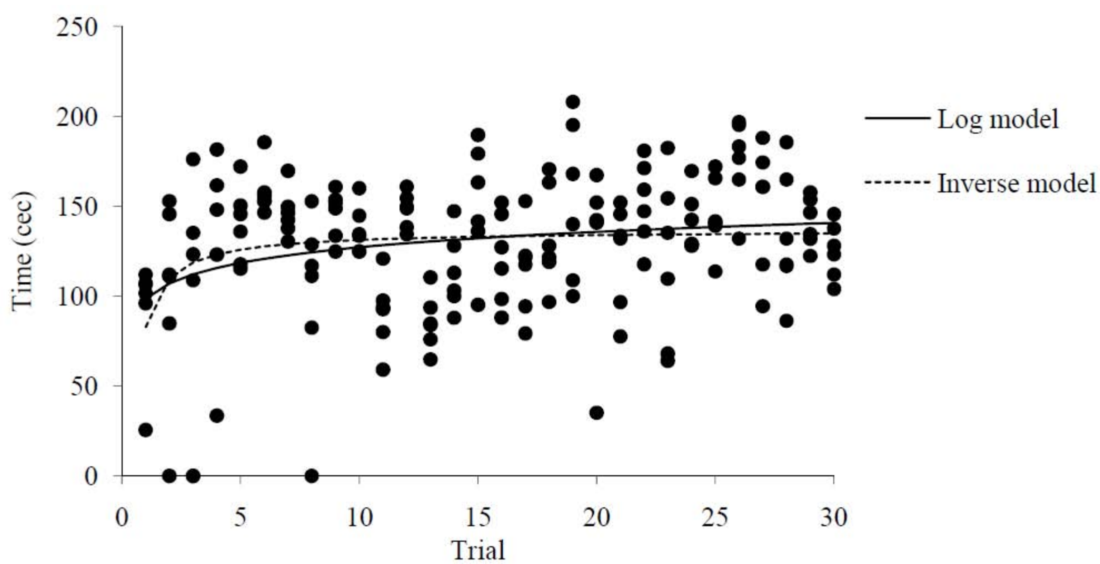


Figure V-19. Scatter plot of the reconnaissance task data for experienced Predator pilots versus the fitted models.

Using our newly defined performance criterion, we converted participants' data on the total time on target per trial for each of 30 trials into the number of trials required until the performance criterion was reached. The response variable of interest, *y*, was

then the proportion of participants in each personnel category who achieved proficiency on the reconnaissance task. A graph of the response variable versus the number of trials is shown in Figure V-20.

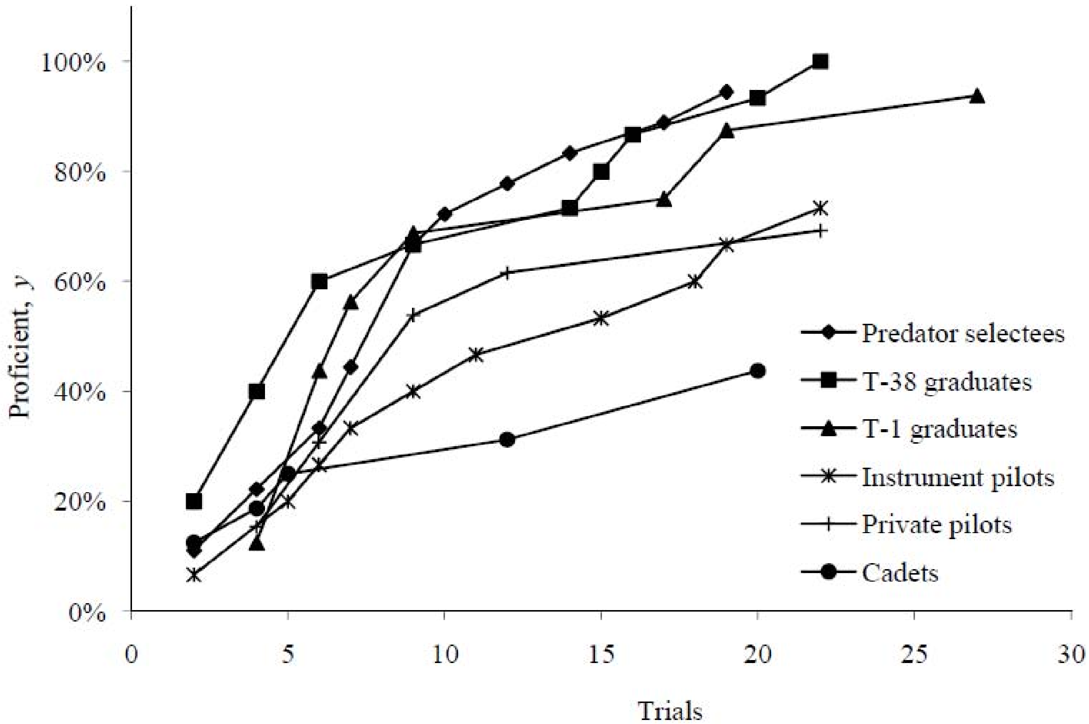


Figure V-20. Scatter plot of the reconnaissance task data.

The independent variables were the continuous variable, *Trial*, corresponding to the number of trials to reach proficiency and the categorical variable, *Group*, corresponding to personnel category. The dependent variable was the proportion proficient, *y*, which was a measure of human performance on the reconnaissance task relative to the derived standard of performance. The basic logistic model related the proportion proficient to the three potential predictor variables, *Trial*, *Group*, and their interaction, $Trial \times Group$. The categorical variable, *Group*, was dummy variable coded for inclusion in the regression analysis. Linear and additive regression models were

examined and plots of the models were used to assess the fit, determine the influence of outliers, and assure regression assumptions were not violated (see the chapter appendix for details).

We found that the first order model with interaction and a logarithmic transformation of the predictor, *Trials*, based on a linear fit and using the logit link function, was the most parsimonious, resulting in the final fitted logistic regression model of:

$$\log\left(\frac{\hat{y}}{1-\hat{y}}\right) = -3.0463 + 0.8110\log(x_1) - 0.5714x_2 - 0.3496x_3 - 1.3588x_4$$
$$- 0.9234x_5 + 0.4258x_6 + 0.6327\log(x_1)x_2 + 0.6604\log(x_1)x_3 \qquad (69)$$
$$+ 1.4636\log(x_1)x_4 + 1.1904\log(x_1)x_5 + 0.7887\log(x_1)x_6$$

where:

$x_1 =$ Trials $[0,+\infty)$

$x_2 =$ Civil instrument pilots $\{0,1\}$

$x_3 =$ Civil private pilots $\{0,1\}$

$x_4 =$ Predator selectees $\{0,1\}$

$x_5 =$ T-1 graduates $\{0,1\}$

$x_6 =$ T-38 graduates $\{0,1\}$

Table V-4 summarizes the estimated regression coefficients and standard errors for the final fitted logistic regression model of the reconnaissance task data. A graph of the fitted response variable $(\hat{y})$ versus number of trials to reach proficiency (x) by personnel category is shown in Figure V-21.

Table V-4. Estimated regression coefficients and standard errors for the final fitted model of the reconnaissance task data.

Variable	$\hat{\beta}$	$\text{se}\left(\hat{\beta}\right)$	Z	p-value
Constant	-3.04627	0.78279	-3.892	<0.0001
Time	0.81102	0.30416	2.666	0.0077
Group(Civil instrument pilots)	-0.57142	1.10229	-0.518	0.6045
Group(Civil private pilots)	-0.34958	1.37933	-0.253	0.8000
Group((Predator selectees)	-1.35876	1.12045	-1.213	0.2251
Group(T-1 graduates)	-0.92344	1.20842	-0.764	0.4449
Group(T-38 graduates)	0.42576	1.02447	0.416	0.6774
Time x Group(Civil instrument pilots)	0.63274	0.43972	1.439	0.1502
Time x Group(Civil private pilots)	0.66037	0.58452	1.130	0.2586
Time x Group(Predator selectees)	1.46360	0.47689	3.069	0.0021
Time x Group(T-1 graduates)	1.19045	0.51613	2.306	0.0211
Time x Group(T-38 graduates)	0.78869	0.42789	1.843	0.0653

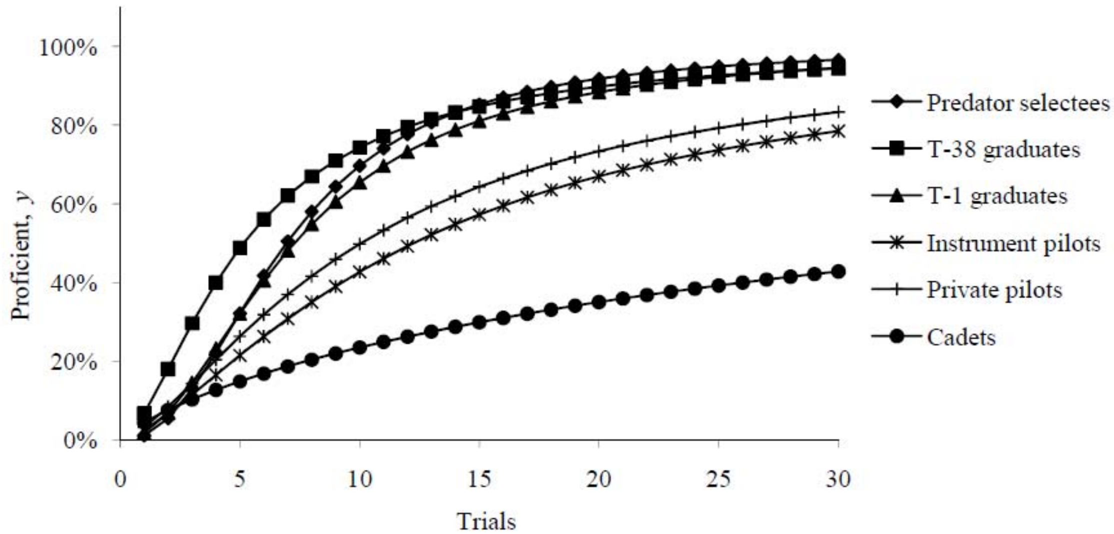


Figure V-21. Plot of the fitted reconnaissance task data.

We next calculated a system of training-reliability equations—one equation for each personnel category—using the coefficient vector and the inverted covariance matrix of model parameters, the latter of which is shown below:

$$\begin{pmatrix}
0.61277 & -0.22526 & -0.61277 & -0.61277 & -0.61277 & -0.61277 & -0.61277 & 0.22526 & 0.22526 & 0.22526 & 0.22526 & 0.22526 \\
-0.22526 & 0.09251 & 0.22526 & 0.22526 & 0.22526 & 0.22526 & 0.22526 & -0.09251 & -0.09251 & -0.09251 & -0.09251 & -0.09251 \\
-0.61277 & 0.22526 & 1.21505 & 0.61277 & 0.61277 & 0.61277 & 0.61277 & -0.46492 & -0.22526 & -0.22526 & -0.22526 & -0.22526 \\
-0.61277 & 0.22526 & 0.61277 & 1.90256 & 0.61277 & 0.61277 & 0.61277 & -0.22526 & -0.77591 & -0.22526 & -0.22526 & -0.22526 \\
-0.61277 & 0.22526 & 0.61277 & 0.61277 & 1.25541 & 0.61277 & 0.61277 & -0.22526 & -0.22526 & -0.51193 & -0.22526 & -0.22526 \\
-0.61277 & 0.22526 & 0.61277 & 0.61277 & 0.61277 & 1.46028 & 0.61277 & -0.22526 & -0.22526 & -0.22526 & -0.59728 & -0.22526 \\
-0.61277 & 0.22526 & 0.61277 & 0.61277 & 0.61277 & 0.61277 & 1.04955 & -0.22526 & -0.22526 & -0.22526 & -0.22526 & -0.41303 \\
0.22526 & -0.09251 & -0.46492 & -0.22526 & -0.22526 & -0.22526 & -0.22526 & 0.19336 & 0.09251 & 0.09251 & 0.09251 & 0.09251 \\
0.22526 & -0.09251 & -0.22526 & -0.77591 & -0.22526 & -0.22526 & -0.22526 & 0.09251 & 0.34166 & 0.09251 & 0.09251 & 0.09251 \\
0.22526 & -0.09251 & -0.22526 & -0.22526 & -0.51193 & -0.22526 & -0.22526 & 0.09251 & 0.09251 & 0.22742 & 0.09251 & 0.09251 \\
0.22526 & -0.09251 & -0.22526 & -0.22526 & -0.22526 & -0.59728 & -0.22526 & 0.09251 & 0.09251 & 0.09251 & 0.26639 & 0.09251 \\
0.22526 & -0.09251 & -0.22526 & -0.22526 & -0.22526 & -0.22526 & -0.41303 & 0.09251 & 0.09251 & 0.09251 & 0.09251 & 0.18309
\end{pmatrix}$$

Fixing the assurance level at 0.90, we then created the following system of equations, noting that we must address the log transformation of x_1:

$$x_1 = \exp\left\{ f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 1\right)\right\} = \exp\left\{4.15297 + 1.60340 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + \sqrt{0.89060 + 1.27272 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 0.59388 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2}\right\} \tag{70}$$

$$x_1 = \exp\left\{ f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 2\right)\right\} = \exp\left\{2.51688 + 0.75242 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + \sqrt{0.02955 + 0.01677 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 0.04498 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2}\right\} \tag{71}$$

$$x_1 = \exp\left\{ f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 3\right)\right\} = \exp\left\{2.33070 + 0.83801 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + \sqrt{0.07084 + 0.03820 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 0.13273 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2}\right\} \tag{72}$$

$$x_1 = \exp\left\{ f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 4\right)\right\} = \exp\left\{1.92817 + 0.45930 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + \sqrt{0.01277 - 0.00774 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 0.00903 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2}\right\} \tag{73}$$

$$x_1 = \exp\left\{ f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 5\right)\right\} = \exp\left\{1.97141 + 0.53798 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + \sqrt{0.02477 - 0.01290 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 0.02063 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2}\right\} \tag{74}$$

$$x_1 = \exp\left\{ f\left(\pi_{\text{spec}} \middle| \alpha = 0.90, k = 6\right)\right\} = \exp\left\{1.61127 + 0.66369 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + \sqrt{0.04477 - 0.03563 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right) + 0.02560 \cdot \log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}}\right)^2}\right\} \tag{75}$$

Again, we traced all combinations of the two determinants, training and personnel, sufficient to provide a specified level of reliability. We set these specifications, given in terms of proportion proficient, equal to 0.50, 0.70, 0.90, 0.95, and 0.99. Figure V-22 provides a graphical display of the resulting isoreliability model for the reconnaissance task data. As with the other tasks, the logistic regression analysis of the reconnaissance task data indicates that personnel category and number of trials (i.e., training) were both significant with training being the more important predictor. However, the isoreliability curves show that the reconnaissance task, as compared to the basic maneuvering and landing tasks, was by far the most sensitive to personnel.

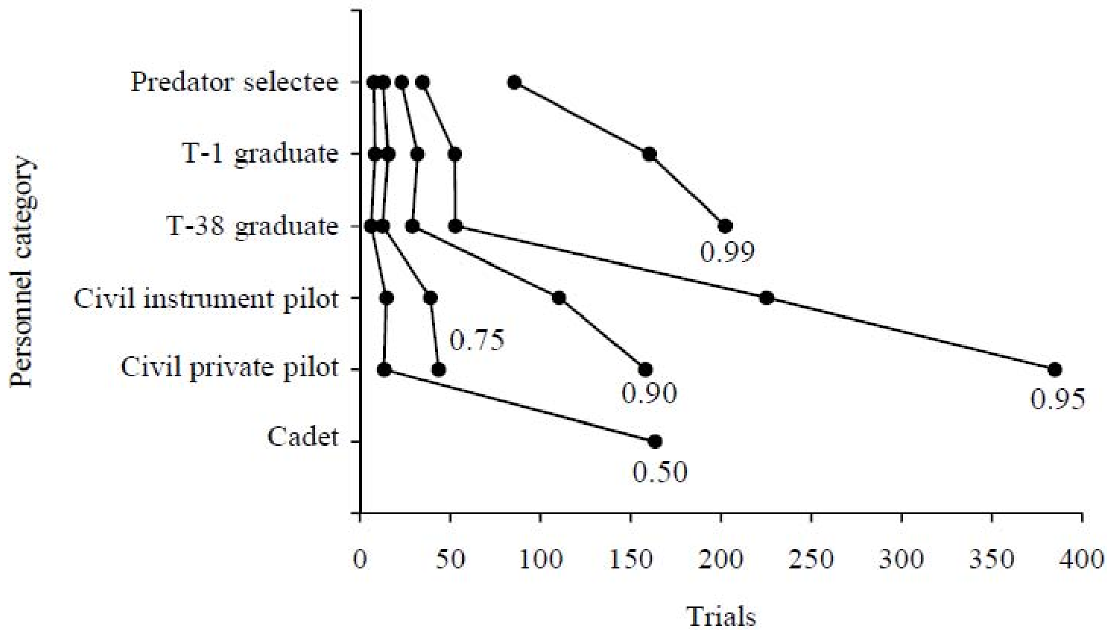


Figure V-22. Isoreliability curves trading off aptitude (personnel category) and training trials with the proportion proficient set at 0.50, 0.75, 0.90, 0.95, and 0.99 and level of assurance set at 0.90 for all criterion settings.

Figure V-23 displays the personnel sensitivity, expressed in terms of the percent difference in training time for each personnel category relative to the Predator selectee category, for various settings of the reliability criterion. In the case of the cadet category, personnel sensitivity was only shown for the 0.50 reliability level; personnel sensitivity was otherwise two to three orders of magnitude greater than that of the next highest category for the remaining reliability levels. For the first time, there is a distinct break

between the set of personnel categories comprised of military pilots and the complementary set formed of civil pilots and non-pilot cadets. The personnel categories corresponding to military pilots appeared to be relatively insensitive, while those corresponding to the non-military pilots exhibited a monotonically increasing trend in personnel sensitivity with increasing reliability criterion levels. This pattern was suggestive of an individual aptitude selected during military pilot screening, a skill gained during military pilot training, or both, that are significant enablers in performing the reconnaissance task.

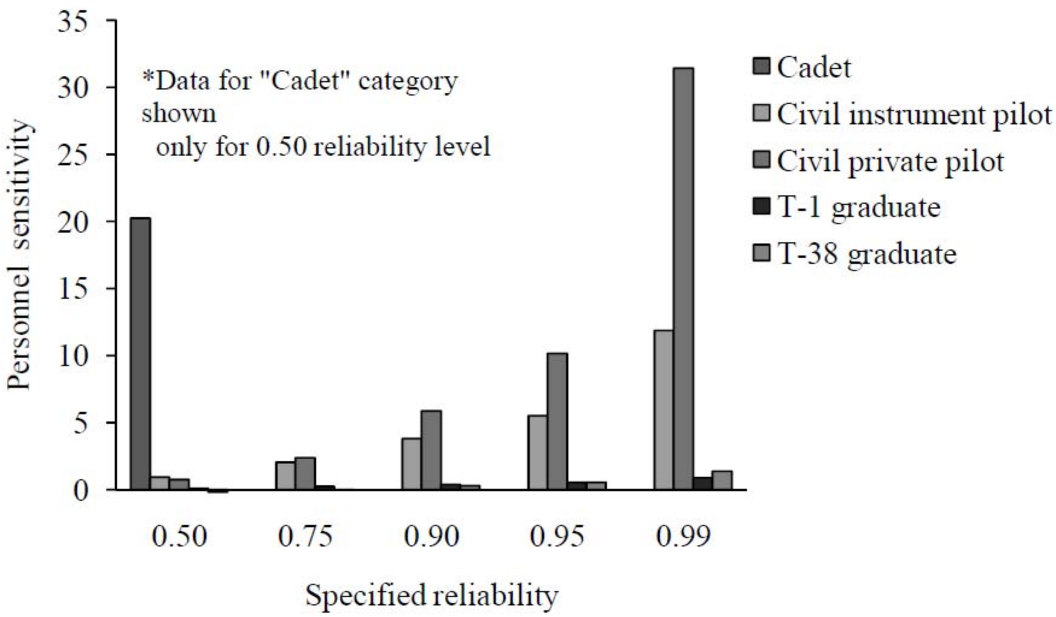


Figure V-23. Personnel sensitivity, expressed in terms of the percent difference in training time for each personnel category relative to the Predator selectee category, for various settings of the reliability criterion on the reconnaissance task.

D. AGGREGATE ISORELIABILITY ANALYSIS

1. Mathematical Programming Formulation

In this section, a method for forming an aggregate isoreliability model was developed. As we discussed earlier, a significant advantage in working with reliability rather than performance is the ability to avail ourselves of basic system models. A

system's functional and physical decomposition can be used to construct a system-level reliability block diagram, the structure of which is used to compute reliability in terms of component and subsystem reliabilities. In the case of our Predator operator, we might consider the reliability block diagram shown in Figure V-24. This diagram was derived, with some adaptation, from a front end analysis of the workflow of a MQ-1 Predator pilot (Nagy, Kalita, & Eaton, 2006). It was simplified such that the functions depicted could be reasonably matched with those tasks assessed by Schreiber and colleagues in their study. Thus, functions 3.2 and 3.4 correspond to the basic maneuvering task, function 3.3 corresponds to the reconnaissance task, and function 3.5 matches up with the landing task. If we assume that functions 3.2 to 3.5 are functionally independent, then the set of functions constitutes a series system. Accordingly, the series system reliability, which is the probability of overall system success, is given by

$$R_s = \prod_{f=3.2}^{3.5} R_{f,e} R_{f,h}\left(x_1, x_2\right) \tag{76}$$

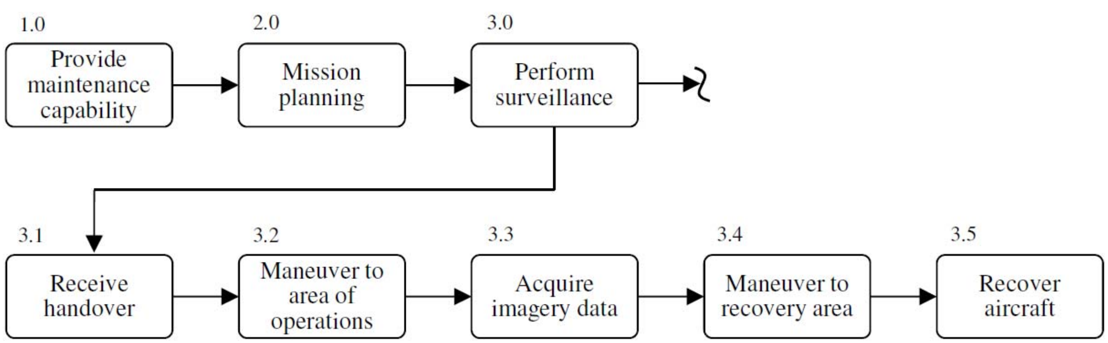


Figure V-24. Reliability block diagram.

A good plan for choosing a Predator operator, particularly from a system sustainability perspective, is to seek a solution that most effectively utilizes personnel given total system reliability and training resource constraints. In such a situation, the quality of feasible solutions then might be judged in terms of maximizing total system reliability for the personnel costs expended. Thus, assigning the indices

f : function $\left(f = 3.2, 3.3, 3.4, 3.5\right)$

p : personnel category $\left(p = \text{Predator selectee}, \ldots, \text{cadet}\right)$

t : task $\left(t = \text{maneuver, landing, recon}\right)$

the following symbolic constants parameterize the objective:

e_f : the equipment reliability for function f

m_p : the personnel costs for an operator from personnel category p

c_t : the hourly cost for training task t

l : the average duration of a landing trial (in minutes).

Then with the binary and nonnegative decision variables

$x_{1,t}$: the amount of training provided for task t

$$x_{2,p} : \begin{cases} 1 \text{ if an operator from personnel category } p \text{ is selected} \\ 0 \text{ otherwise} \end{cases}$$

$h_{f,p}$: carrier variable for the human reliability value for function f given personnel category p

w_f : carrier variable for amount of training provided for function f

y_t : carrier variable for training time (in hours) for task t

we have the following objective:

$$\max \quad \frac{\prod_f \left[e_f \left(\sum_p h_{f,p} x_{2,p} \right) \right]}{\sum_p m_p x_{2,p} + \sum_t c_t y_t} \tag{77}$$

The numerator simply reduces to Equation 76 once a personnel category is chosen.

To complete the model, we must enforce several constraints. Again, we define symbolic parameters

$\underline{r}$: the lower limit on acceptable total system reliability

$\overline{t}$: the upper limit on available training time (in minutes)

The first system of constraints relates the amount of training provided for task t as defined in the study by Schreiber and colleagues to the corresponding function(s) f as depicted in Figure V-24:

$$
\begin{aligned}
&w_{3.2} = w_{3.4} = x_{1,\text{maneuver}} \\
&w_{3.3} = x_{1,\text{recon}} \\
&w_{3.5} = x_{1,\text{landing}}
\end{aligned} \tag{78}
$$

The second system of constraints relates the personnel and training domains to functional reliability using our empirically derived isoreliability curves:

$$
h_{f,p} = g_{f,p}\left(w_f\right) \quad \forall\, f, p \tag{79}
$$

Our third constraint is a partitioning constraint requiring that exactly one personnel category be chosen in the solution:

$$
\sum_p x_{2,p} = 1. \tag{80}
$$

The fourth constraint enforces the lower limit on total system reliability given a choice of personnel category:

$$
\underline{r} \leq \prod_f \left[e_f \left(\sum_p h_{f,p} x_{2,p} \right) \right] \tag{81}
$$

The next system of constraints simply relates the amount of training for each task to the equivalent training time in hours:

$$
\begin{aligned}
&y_{\text{maneuver}} = \frac{x_{1,\text{maneuver}}}{60} \\
&y_{\text{landing}} = \frac{l x_{1,\text{landing}}}{60} \\
&y_{\text{recon}} = \frac{10 x_{1,\text{recon}}}{60}
\end{aligned} \tag{82}
$$

Our last constraint enforces the upper limit on training time:

$$
\overline{t} \geq \sum_t y_t \tag{83}
$$

For the sake of tractability, we formulate the system of equations indicated by $g_{f,p}\left(w_f\right)$ starting directly from Equation 46:

$$
\log\left(\frac{\pi_{\text{spec}}}{1-\pi_{\text{spec}}} \right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_k + \hat{\beta}_{k+5} x_1 - z_\alpha \sqrt{c_{1k} + c_{2k} x_1 + c_{3k} x_1^2}
$$

It may help to recall from our earlier derivation of Equation 46 that π_{spec} is the specified reliability, x_1 is the amount of training provided, the βs are regression coefficients, and k indexes personnel category such that $k = \{2,3,4,5,6\}$. Additionally, c_{1k}, c_{2k}, and c_{3k} are constants calculated from elements of the regression covariance matrix according to Equation 44. It should also be recalled that in the special case where $k = 1$, there are no $\hat{\beta}_k$ or $\hat{\beta}_{k+5}$ terms in Equation 46 and $c_{1,1}$, $c_{2,1}$, and $c_{3,1}$ can be inferred from Equation 43. However, we will fold this special case into the more generic formulation shortly.

As we are no longer fixing $\pi = \pi_{\text{spec}}$, we can rewrite Equation 46 so that we simply have the variable π on the left hand side:

$$\pi = \frac{1}{1+\mathrm{e}^{-d}} \tag{84}$$

where

$$d = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_k + \hat{\beta}_{k+5} x_1 - z_\alpha \sqrt{c_{1k} + c_{2k} x_1 + c_{3k} x_1^2} \tag{85}$$

It should be obvious that we will have a unique equation for d for each personnel category, and so we can think of Equations 84-85 as actually being a family of personnel category-specific equations. Accordingly, we can re-index Equations 84-85 in terms of p and directly include all k personnel categories by stipulating that $\beta_k = \beta_{k+5} = 0$ when $k = 1$, thereby eliminating the need to track a special case:

$$\pi_p = \frac{1}{1+\mathrm{e}^{-d}} \tag{86}$$

where

$$d = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_{2p} + \hat{\beta}_{3p} x_1 - z_\alpha \sqrt{c_{1p} + c_{2p} x_1 + c_{3p} x_1^2} \tag{87}$$

It should also be noted that the regression coefficients are specific to the model fitted for data from some task, *t*, which corresponds to one or more functions, *f*. Thus, it is also possible to specify a unique equation for *d* for each function, *f*:

$$\pi_{f,p} = \frac{1}{1+\mathrm{e}^{-d}} \tag{88}$$

where

$$d = \hat{\beta}_{0f} + \hat{\beta}_{1f} x_1 + \hat{\beta}_{2fp} + \hat{\beta}_{3fp} x_1 - z_\alpha \sqrt{c_{1fp} + c_{2fp} x_1 + c_{3fp} x_1^2} \tag{89}$$

Examining Equations 88-89, it should be evident that, with the exception of x_1, all the terms on the right hand side of the equation are constants: their values are obtained from the corresponding vector of estimated regression coefficients or the covariance matrix of model parameters. In terms of our formulation of an aggregate isoreliability model, they are simply elements stored in a data matrix, **D**. Consequently:

$$h_{f,p} = \pi_{f,p} = g_{f,p}\left(w_f\right) = \frac{1}{1+\exp\left\{-\left(d_{1,f,p} + d_{2,f,p} w_f + d_{3,f,p} + d_{4,f,p} w_f - z_\alpha \sqrt{d_{5,f,p} + d_{6,f,p} w_f + d_{7,f,p} w_f^2}\right)\right\}} \tag{90}$$

where we now denote the amount of training provided for function, *f*, using the variable w_f rather than the more generic x_1.

2. Example

The Predator operator selection problem is idealized to consider just functions 3.2 to 3.5 as shown in Figure V-24. The overall objective is to maximize (larger-is-better) cost effectiveness in terms of system reliability relative to personnel costs. The reliability of the equipment, e_f, is taken as a given in this case, meaning $e_f = 1 \; \forall \; f$. Reasonable specifications that might be expressed by a decision maker are as follows: 1) total system reliability must be no less than 0.90 $\left(\underline{r} = 0.90\right)$, and 2) simulator availability per trainee is limited to no more than 40 hours $\left(\overline{t} = 40\right)$. With regards to the training simulator, the average time for a landing trial is 5 minutes $\left(l = 5\right)$. Estimates for personnel costs, m_p, are summarized in Table V-5. Hourly simulator rates, c_t, are estimated at \$40 per hour for the maneuvering and landing tasks and \$65 per hour for the reconnaissance task. These rates are derived from the fee schedule for the light aircraft simulator used in Kansas State University's aviation training program. The rates for the maneuvering and landing tasks assume solo training, while that for the reconnaissance task assumes dual instruction with an advanced flight instructor. The choice of different rates reflects the

fact that explicit performance feedback is available for the maneuvering and landing tasks in the synthetic task environment used by Schreiber and colleagues, while the reconnaissance task is more freeform. These rates were normalized as per the personnel costs in Table V-5: $c_{\text{maneuver}} = c_{\text{landing}} = 0.004$ and $c_{\text{recon}} = 0.006$.

Table V-5. Personnel cost estimates.

Personnel category	Manpower cost elements	Estimates (FY05)	Normalized costs[8]
Predator selectee	SMCR[1] O-3	$100,833[5]	73.783
	SUPT[2]	392,861[7]	
	B-52 IQT[3]	292,190[7]	
	Total	$785,934	
T-38 graduate	SMCR[1] O-1	$ 62,982[5]	42.794
	SUPT[2]	392,861[7]	
	Total	$455,843	
T-1 graduate	SMCR[1] O-1	$ 62,982[5]	42.794
	SUPT[2]	392,861[7]	
	Total	$455,843	
Civil instrument pilot	SMCR[1] O-1	$ 62,982[5]	7.039
	IFT[4]	5,500[7]	
	Instrument rating	6,500[7]	
	Total	$74,982	
Civil private pilot	SMCR[1] O-1	$ 62,982[5]	6.429
	IFT[4]	5,500[7]	
	Total	$68,482	
Cadet	SMCR[1] Cadet	$10,652[6]	1.000

[1]SMCR = Standard military compensation rate
[2]SUPT = Specialized undergraduate pilot training
[3]IQT = Initial qualification training
[4]IFT = Initial flight training
[5]Source: Dahlman, 2005
[6]Source: DoD Comptroller, 2010
[7]Source: Hoffman & Kamps, 2005
[8]Relative to cadet

Below is the nonlinear integer program for our Predator operator selection problem. To simplify notation, we use the following abbreviations in the indexing for personnel category: S = Predator selectee, T38 = T-38 graduate, T1 = T-1 graduate, I = civil instrument pilot, P = civil private pilot, C = cadet. Similarly, we use the following abbreviations in the indexing for task: M = maneuver, L = landing and R = reconnaissance.

$$\max \quad \frac{\begin{bmatrix} \left(h_{3.2,\mathrm{S}}x_{2,\mathrm{S}} + h_{3.2,\mathrm{T38}}x_{2,\mathrm{T38}} + h_{3.2,\mathrm{T1}}x_{2,\mathrm{T1}} + h_{3.2,\mathrm{I}}x_{2,\mathrm{I}} + h_{3.2,\mathrm{P}}x_{2,\mathrm{P}} + h_{3.2,\mathrm{C}}x_{2,\mathrm{C}}\right)\cdot \\ \left(h_{3.3,\mathrm{S}}x_{2,\mathrm{S}} + h_{3.3,\mathrm{T38}}x_{2,\mathrm{T38}} + h_{3.3,\mathrm{T1}}x_{2,\mathrm{T1}} + h_{3.3,\mathrm{I}}x_{2,\mathrm{I}} + h_{3.3,\mathrm{P}}x_{2,\mathrm{P}} + h_{3.3,\mathrm{C}}x_{2,\mathrm{C}}\right)\cdot \\ \left(h_{3.4,\mathrm{S}}x_{2,\mathrm{S}} + h_{3.4,\mathrm{T38}}x_{2,\mathrm{T38}} + h_{3.4,\mathrm{T1}}x_{2,\mathrm{T1}} + h_{3.4,\mathrm{I}}x_{2,\mathrm{I}} + h_{3.4,\mathrm{P}}x_{2,\mathrm{P}} + h_{3.4,\mathrm{C}}x_{2,\mathrm{C}}\right)\cdot \\ \left(h_{3.5,\mathrm{S}}x_{2,\mathrm{S}} + h_{3.5,\mathrm{T38}}x_{2,\mathrm{T38}} + h_{3.5,\mathrm{T1}}x_{2,\mathrm{T1}} + h_{3.5,\mathrm{I}}x_{2,\mathrm{I}} + h_{3.5,\mathrm{P}}x_{2,\mathrm{P}} + h_{3.5,\mathrm{C}}x_{2,\mathrm{C}}\right) \end{bmatrix}}{\begin{pmatrix} 73.783x_{2,\mathrm{S}} + 42.794x_{2,\mathrm{T38}} + 42.794x_{2,\mathrm{T1}} + 7.039x_{2,\mathrm{I}} + 6.429x_{2,\mathrm{P}} + x_{2,\mathrm{C}} \\ +0.004y_{\mathrm{M}} + 0.004y_{\mathrm{L}} + 0.006y_{\mathrm{R}} \end{pmatrix}}$$

s.t. $w_{3.2} = w_{3.4} = x_{1,\mathrm{M}}$

$$w_{3.3} = x_{1,\mathrm{R}}$$

$$w_{3.5} = x_{1,\mathrm{L}}$$

$$h_{3.2,\mathrm{S}} = \frac{1}{1 + \mathrm{e}^{3.7966 - 0.0708w_{3.2} + 3.7123 - 0.0998w_{3.2} + 1.2816\sqrt{0.9093 - 0.0391w_{3.2} + 0.0004w_{3.2}^2}}}$$

$$h_{3.2,\mathrm{T38}} = \frac{1}{1 + \mathrm{e}^{3.7966 - 0.0708w_{3.2} + 0.7665 - 0.0253w_{3.2} + 1.2816\sqrt{0.4546 - 0.0177w_{3.2} + 0.0002w_{3.2}^2}}}$$

$$h_{3.2,\mathrm{T1}} = \frac{1}{1 + \mathrm{e}^{3.7966 - 0.0708w_{3.2} + 1.4358 - 0.0318w_{3.2} + 1.2816\sqrt{0.5211 - 0.0198w_{3.2} + 0.0002w_{3.2}^2}}}$$

$$h_{3.2,\mathrm{I}} = \frac{1}{1 + \mathrm{e}^{3.7966 - 0.0708w_{3.2} + 1.6245 - 0.0555w_{3.2} + 1.2816\sqrt{0.6236 - 0.0277w_{3.2} + 0.0003w_{3.2}^2}}}$$

$$h_{3.2,\mathrm{P}} = \frac{1}{1 + \mathrm{e}^{3.7966 - 0.0708w_{3.2} + 1.8577 - 0.0412w_{3.2} + 1.2816\sqrt{1.0601 - 0.0414w_{3.2} + 0.0004w_{3.2}^2}}}$$

$$h_{3.2,\mathrm{C}} = \frac{1}{1 + \mathrm{e}^{3.7966 - 0.0708w_{3.2} + 1.2816\sqrt{0.3371 - 0.0123w_{3.2} + 0.0001w_{3.2}^2}}}$$

$$h_{3.3,\mathrm{S}} = \frac{1}{1+e^{3.0463-0.8110\log(w_{3.3})+1.3588-1.4636\log(w_{3.3})+1.2816\sqrt{0.6426-0.5733\log(w_{3.3})+0.1349(\log(w_{3.3}))^2}}}$$

$$h_{3.3,\mathrm{T38}} = \frac{1}{1+e^{3.0463-0.8110\log(w_{3.3})-0.4258-0.7887\log(w_{3.3})+1.2816\sqrt{0.4368-0.3755\log(w_{3.3})+0.0906(\log(w_{3.3}))^2}}}$$

$$h_{3.3,\mathrm{T1}} = \frac{1}{1+e^{3.0463-0.8110\log(w_{3.3})+0.9234-1.1904\log(w_{3.3})+1.2816\sqrt{0.8475-0.7440\log(w_{3.3})+0.1739(\log(w_{3.3}))^2}}}$$

$$h_{3.3,\mathrm{I}} = \frac{1}{1+e^{3.0463-0.8110\log(w_{3.3})+0.5714-0.6327\log(w_{3.3})+1.2816\sqrt{0.6023-0.4793\log(w_{3.3})+0.1008(\log(w_{3.3}))^2}}}$$

$$h_{3.3,\mathrm{P}} = \frac{1}{1+e^{3.0463-0.8110\log(w_{3.3})+0.3496-0.6604\log(w_{3.3})+1.2816\sqrt{1.2898-1.1013\log(w_{3.3})+0.2491(\log(w_{3.3}))^2}}}$$

$$h_{3.3,\mathrm{C}} = \frac{1}{1+e^{3.0463-0.8110\log(w_{3.3})+1.2816\sqrt{0.6128-0.4505\log(w_{3.3})+0.0925(\log(w_{3.3}))^2}}}$$

$$h_{3.4,\mathrm{S}} = \frac{1}{1+e^{3.7966-0.0708w_{3.4}+3.7123-0.0998w_{3.4}+1.2816\sqrt{0.9093-0.0391w_{3.4}+0.0004w_{3.4}^2}}}$$

$$h_{3.4,\mathrm{T38}} = \frac{1}{1+e^{3.7966-0.0708w_{3.4}+0.7665-0.0253w_{3.4}+1.2816\sqrt{0.4546-0.0177w_{3.4}+0.0002w_{3.4}^2}}}$$

$$h_{3.4,\mathrm{T1}} = \frac{1}{1+e^{3.7966-0.0708w_{3.4}+1.4358-0.0318w_{3.4}+1.2816\sqrt{0.5211-0.0198w_{3.4}+0.0002w_{3.4}^2}}}$$

$$h_{3.4,\mathrm{I}} = \frac{1}{1+e^{3.7966-0.0708w_{3.4}+1.6245-0.0555w_{3.4}+1.2816\sqrt{0.6236-0.0277w_{3.4}+0.0003w_{3.4}^2}}}$$

$$h_{3.4,\mathrm{P}} = \frac{1}{1+e^{3.7966-0.0708w_{3.4}+1.8577-0.0412w_{3.4}+1.2816\sqrt{1.0601-0.0414w_{3.4}+0.0004w_{3.4}^2}}}$$

$$h_{3.4,\mathrm{C}} = \frac{1}{1+e^{3.7966-0.0708w_{3.4}+1.2816\sqrt{0.3371-0.0123w_{3.4}+0.0001w_{3.4}^2}}}$$

$$h_{3.5,\mathrm{S}} = \frac{1}{1+e^{3.5976-0.0325w_{3.5}+0.4717-0.0531w_{3.5}+1.2816\sqrt{0.2675-0.0106w_{3.5}+0.0001w_{3.5}^2}}}$$

$$h_{3.5,\mathrm{T38}} = \frac{1}{1+e^{3.5976-0.0325w_{3.5}-0.7784-0.0200w_{3.5}+1.2816\sqrt{0.1815-0.0059w_{3.5}+0.0001w_{3.5}^2}}}$$

$$h_{3.5,\mathrm{T1}} = \frac{1}{1+e^{3.5976-0.0325w_{3.5}-1.5178+0.0054w_{3.5}+1.2816\sqrt{0.0924-0.0019w_{3.5}+0.00001w_{3.5}^2}}}$$

$$h_{3.5,\mathrm{I}} = \frac{1}{1+e^{3.5976-0.0325w_{3.5}-0.8351-0.0317w_{3.5}+1.2816\sqrt{0.3139-0.0132w_{3.5}+0.0002w_{3.5}^2}}}$$

$$h_{3.5,\mathrm{P}} = \frac{1}{1+e^{3.5976-0.0325w_{3.5}-1.8278+0.0025w_{3.5}+1.2816\sqrt{0.1085-0.0026w_{3.5}+0.00002w_{3.5}^2}}}$$

$$h_{3.5,\mathrm{C}} = \frac{1}{1+e^{3.5976-0.0325 w_{3.5}+1.2816\sqrt{0.2880-0.0050 w_{3.5}+0.00002 w_{3.5}^2}}}$$

$$x_{2,\mathrm{S}} + x_{2,\mathrm{T38}} + x_{2,\mathrm{T1}} + x_{2,\mathrm{I}} + x_{2,\mathrm{P}} + x_{2,\mathrm{C}} = 1$$

$$\begin{aligned} 0.90 \le &\left(h_{3.2,\mathrm{S}}x_{2,\mathrm{S}} + h_{3.2,\mathrm{T38}}x_{2,\mathrm{T38}} + h_{3.2,\mathrm{T1}}x_{2,\mathrm{T1}} + h_{3.2,\mathrm{I}}x_{2,\mathrm{I}} + h_{3.2,\mathrm{P}}x_{2,\mathrm{P}} + h_{3.2,\mathrm{C}}x_{2,\mathrm{C}}\right)\cdot \\ &\left(h_{3.3,\mathrm{S}}x_{2,\mathrm{S}} + h_{3.3,\mathrm{T38}}x_{2,\mathrm{T38}} + h_{3.3,\mathrm{T1}}x_{2,\mathrm{T1}} + h_{3.3,\mathrm{I}}x_{2,\mathrm{I}} + h_{3.3,\mathrm{P}}x_{2,\mathrm{P}} + h_{3.3,\mathrm{C}}x_{2,\mathrm{C}}\right)\cdot \\ &\left(h_{3.4,\mathrm{S}}x_{2,\mathrm{S}} + h_{3.4,\mathrm{T38}}x_{2,\mathrm{T38}} + h_{3.4,\mathrm{T1}}x_{2,\mathrm{T1}} + h_{3.4,\mathrm{I}}x_{2,\mathrm{I}} + h_{3.4,\mathrm{P}}x_{2,\mathrm{P}} + h_{3.4,\mathrm{C}}x_{2,\mathrm{C}}\right)\cdot \\ &\left(h_{3.5,\mathrm{S}}x_{2,\mathrm{S}} + h_{3.5,\mathrm{T38}}x_{2,\mathrm{T38}} + h_{3.5,\mathrm{T1}}x_{2,\mathrm{T1}} + h_{3.5,\mathrm{I}}x_{2,\mathrm{I}} + h_{3.5,\mathrm{P}}x_{2,\mathrm{P}} + h_{3.5,\mathrm{C}}x_{2,\mathrm{C}}\right) \end{aligned}$$

$$60 y_{\mathrm{M}} = x_{1,\mathrm{M}}$$

$$12 y_{\mathrm{L}} = x_{1,\mathrm{L}}$$

$$6 y_{\mathrm{R}} = x_{1,\mathrm{R}}$$

$$40 \ge y_{\mathrm{M}} + y_{\mathrm{L}} + y_{\mathrm{R}}$$

The optimal solution has the following applicable non-zero variables:

$$\begin{array}{llll} x_{1,\mathrm{M}}^* = 153.4 & w_{3.2}^* = 153.4 & \mathrm{y}_{\mathrm{M}}^* = 2.6 & h_{3.2,\mathrm{I}}^* = 1.0000 \\ x_{1,\mathrm{L}}^* = 163.1 & w_{3.3}^* = 143.1 & \mathrm{y}_{\mathrm{L}}^* = 13.6 & h_{3.3,\mathrm{I}}^* = 0.9220 \\ x_{1,\mathrm{R}}^* = 143.1 & w_{3.4}^* = 153.4 & \mathrm{y}_{\mathrm{R}}^* = 23.8 & h_{3.4,\mathrm{I}}^* = 1.0000 \\ x_{2,\mathrm{I}}^* = 1 & w_{3.5}^* = 163.1 & & h_{3.5,\mathrm{I}}^* = 0.9970 \end{array}:$$

Figure V-25 shows the plot of cost versus total reliability for alternative feasible solutions, with the maximum attainable objective values for cost effectiveness given in parentheses. The solution using T-1 graduates is dominated by the solution using T-38 graduates, leaving only three solutions comprising an efficient set with regards to the two major elements of the objective function: civilian instrument pilot, T-38 graduate, and Predator selectee. The optimal solution in terms of cost effectiveness is to select civilian instrument pilots and provide 2.6 hours of simulator training for the maneuver task, 13.6 hours of training for the landing task, and 23.8 hours of training for the reconnaissance task. The resulting functional reliabilities are 1.0000 for maneuver (i.e., functions 3.2 and 3.4), 0.9220 for reconnaissance (i.e., function 3.3), and 0.9970 for landing (i.e., function 3.5), thereby achieving a total system reliability of 0.9192.

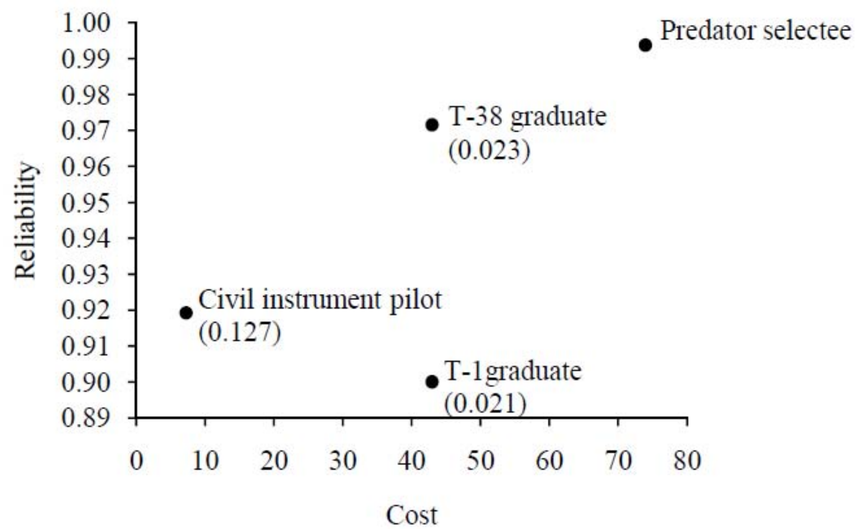


Figure V-25. Cost versus total reliability plot of feasible solutions for the Predator operator selection problem (cost effectiveness objective values given in parentheses).

E. DISCUSSION

Dr. John Weisz (1967), an early pioneer in the field of HSI, advocated that human factors* contributions should be included directly among those factors considered in the "system analytic thinking process" and not treated separately. By system analytic thinking process, Weisz was likely referring to the RAND style systems analysis that was being transferred to the Defense Department by then Secretary of Defense Robert McNamara. Thus, Weisz intended for the inclusion of human factors contributions in analyses comparing, contrasting, and evaluating proposed system concepts, particularly when dealing with weapon systems. Moreover, such analyses were likely to use the quantitative techniques being advanced by the proponents of operations research (Hughes, 1998). Weisz (1967) noted that the challenge for the human factors field was to transfer the research being conducted in the behavioral sciences and human factors engineering fields utilizing experimental psychology methodologies and multi-variable

* The term "HSI" did not come into the Defense Department lexicon until the early 1990s. At the time that Weisz was writing about incorporating human factors considerations in system analyses, he was also advocating for a more broad interpretation of human factors to include those areas that would eventually comprise the U.S. Army's MANPRINT program, the forerunner of the Defense Department's HSI program. Thus, it is reasonable to consider human factors and HSI as synonyms in Weisz's writings.

statistical techniques into the operations research techniques employed in weapon system analysis. He subsequently called out four areas for which human factors could plausibly make contributions to systems analysis: manpower requirements, training requirements, performance requirements, and system design. Ideally, in this analytic paradigm, the human factors team could show tradeoff results between these subcomponents of the human component(s) of the overall system (Weisz, 1968).

This study of Predator UAS personnel requirements directly answers Weisz's challenge, taking an experiment from the behavioral sciences field and transferring the results into mathematical models that lend themselves to analysis utilizing the optimization techniques of operations research. The resulting system analysis collectively considers manpower, training, and performance requirements in selecting a personnel category, allocating task training times, and determining the resulting system reliability and cost effectiveness. While it would have been possible to also consider system design in this system analysis, it was not addressed because proposed system concepts only involved changes in the human component of the system and not the human-machine interface technology. Overall, the study clearly illustrates a general framework around which a detailed model can be developed to permit HSI factors to be effectively introduced into system analysis studies.

Besides allowing HSI factors to be considered in the initial system concept comparison, the underlying mathematical models can be directly used by analysts to assist decision makers who may want to conduct tradeoff studies. This process allows the HSI team to show tradeoff results between domain-related decision variables that contribute to human performance. It is very likely, for example, that decreasing personnel quality will decrease personnel-related costs but necessitate an increase in one or more elements of training, thereby increasing support costs. Additionally, total-system analysts could perform tradeoffs between major areas contributing to the system analysis, for example by changing equipment reliabilities or elements of the reliability block diagram itself, until some optimum combination of all areas is achieved. Furthermore, the optimum combination can be sought for wartime conditions when the cost of training

and, especially, the time available for such training on a mass basis may make certain personnel and training solutions infeasible (Weisz, 1968).

As mentioned earlier, the primary challenge to bringing behavioral science and human factors engineering experimental results into system analyses is the necessity of formulating the results of such experiments as mathematical models that are tractable to the techniques of operations research. The isoperformance methodology described by Jones, Kennedy, and their colleagues provides an intuitively simple but highly useful approach for making such a transfer. In particular, the motivation for suggesting isoperformance is to make HSI domain tradeoff decisions more mathematically tractable, and hence, explicit—an idea that should clearly resonate with system analysis practitioners. However, their demonstrations of the isoperformance method are limited to single-function cases with continuous decision variables. The reality, as typified by the Predator operator selection problem, is that many critical decision problems in engineering and management include logical decisions that cannot be modeled validly as continuous and/or are concerned with performance of aggregates of functions, which is to say systems. Thus, an important finding of this study is the feasibility of 1) including logical decision variables in isoperformance models and 2) the ability to transfer those models into discrete optimization models that can then be used to analyze aggregated functions, at least in terms of the construct of human reliability. It should also come as no surprise that in so doing, we necessarily increase the complexity of problem formulation and diminish the tractability of the resulting optimization models.

While we have used the term "human reliability" as if it were a novel discovery, the truth is that interest in human reliability can be traced back to the middle 1950s, along with, or perhaps as one aspect of, the then growing interest in systems theory. As an activity, human reliability involves the quantitative analysis, prediction, and evaluation of procedural, goal-directed human performance in human-machine systems in such terms as error likelihood, probability of task accomplishment, and response time (Meister, 1985). As described in Pew's (2008) 50-year retrospect on human performance model development, Swain and his group at Sandia National Laboratories were major innovators in this area, creating the well known technique for human error rate

prediction, or simply THERP. While there are other tools for human reliability analysis and prediction, such as Siegel/Wolf models, which are the basis for IMPRINT[21], we will confine ourselves chiefly with THERP as the archetype for these methods. Briefly, THERP recognizes that performing typical human tasks involves the serial aggregation of collections of elemental actions, and as task analysis reveals, the aggregation involves a contingent branching structure of possible paths through a network of such actions. Accordingly, THERP decomposes human tasks into their elemental actions, assigns probabilities of error for each element, and through serial aggregation, calculates the probability of error for a task (Pew, 2008).

A number of factors complicate this apparently straightforward procedure, not the least of which is the determination of possible error rates and relative influence of factors that affect error likelihood, the latter being called *performance shaping factors*. A primary consideration in conducting a human reliability analysis is the variability of the performance of interest. Because of this variability, the reliability of human performance usually is not predicted solely as a point estimate, but is considered to lie within a range of uncertainty. To arrive at a point estimate for human error probability, a point value is chosen from within a given range based on an analyst's assessment of the impact of performance shaping factors. For example, if an analyst believes a performance-shaping factor exists that will increase the likelihood of error, they would select a human error probability closer to the upper bound of the uncertainty range for the nominal value (Meister, 1985). In the case of Siegel/Wolf models, the issue of performance variability is handled by creating Monte Carlo simulations of reliability networks and performance-shaping factors are implemented as scale factors (Meister, 1985; Pew, 2008).

[21] The U.S. Air Force used Siegel's modeling approach to develop SAINT (Systems Analysis of Integrated Networks of Tasks), which was a general-purpose discrete simulation language written in FORTRAN. Micro Analysis and Design (MAAD) subsequently developed a PC version in 1986 (Micro Saint) written in C, which then spawned a family of special-purpose applications. The most prominent thread in this lineage is contained in the IMPRINT series of applications, mostly sponsored by the U.S. Army, which is widely used to model operator workload and task performance. A major thrust in the evolution of IMPRINT has been the modeling of the dynamic relationship between mental workload, as predicted based on Wickens' Multiple Resource Theory, and performance (McCracken & Aldrich, 1984; Mitchell, 2000; Pew, 2008).

HSI practitioners looking at THERP (or Siegel/Wolf models for that matter) would likely nominate personnel and training domain considerations as being key, modifiable performance shaping factors. However, when using THERP, the impact of these factors on system reliability is largely left to the judgment of the analyst performing the human reliability analysis. Tradeoffs between performance shaping factors must be addressed by selecting various values for a given parameter from the distribution of error probabilities available and modifying these by some aggregate performance shaping "fudge" factor—a clumsy process at best. The task simulation method, made possible by Siegel/Wolf models, has the advantage that one can experimentally vary performance shaping factor levels to determine their effect on system reliability, but the validity of the results will obviously depend on the correctness of the algorithms that make up the model architecture (i.e., how accurately the performance shaping factors and their interaction are implemented). Moreover, in the end, such simulations are little more than a source of data and it still remains for an analyst to derive some mathematical tradeoff function linking performance shaping factor levels to organizational objectives. Thus, the relative contribution of our adaptation of Jones and Kennedy's isoperformance methodology to human reliability analysis should be evident, particularly when an analyst is seeking to explicitly include HSI domain considerations within larger system analyses.

Returning to the case of the Predator UAS, personnel quality issues were not addressed early in the acquisition process because it was simply assumed at the outset that experienced, rated Air Force aviators would be the system operators. Only as the growth in the number of systems outpaced the available personnel resources has the Air Force recognized the necessity to reconsider personnel quality requirements for the Predator UAS. Necessarily limiting ourselves to the *ex post facto* data generated by Schreiber and colleagues, the analyses performed in this study indicate that some less capable personnel may be considered for the Predator operator job while still meeting or exceeding the desired level of total system reliability. As was shown in the tradeoff analysis, experienced Air Force aviators are still an ideal choice in terms of overall system reliability, but their associated personnel costs make them a sub-optimal choice with regards to cost effectiveness. A similar statement can be made with regards to rated

aviators graduating from the fighter/bomber track of specialized undergraduate pilot training. Therefore, the use of rated Air Force aviators as Predator UAS operators might be construed as an example of "gold plating," a term that is derisively used more often with regards to technical system components to convey the notion of excess capability.

An analysis, similar to that performed in this study, could have estimated these personnel quality requirements early in the acquisition process of this weapon system. These analyses might well have been employed to more cost effectively achieve system reliability requirements by either designing the equipment to be less complex, decreasing standards for personnel capability, or by a combination of both. Addressing system reliability requirements via the equipment would require the engineers to redesign the system to reduce the complexity of the tasks that are performed by the operators. Using this method often leads to the equipment becoming more complex internally, and hence, more expensive.

Care must also be taken not to diminish overall system reliability by decreasing equipment reliability for gains in human reliability. Moreover, it is very important to appreciate the relative aptitude sensitivity of those tasks and functions allocated to the system operator so that efforts to reduce task complexity are appropriately targeted. This last point is well illustrated by the isoplots constructed in this study, which clearly show the reconnaissance task to be the major limitation when attempting to relax personnel capability requirements. Given ever present financial resource constraints, efforts to reduce task complexity might be better focused on decision aiding during the reconnaissance task rather than fielding an automatic landing system. Again, such system concepts should be subjected to some type of system analysis so that a more optimal solution is chosen.

It should be noted that the tasks analyzed by Schreiber and colleagues are not representative of all the tasks and functions relevant to operation of the Predator UAS. Nor did the cognitive task analysis on which Martin and Schreiber (1998) based the development of their synthetic task environment consider the mission of *armed* reconnaissance as this was not part of the Predator mission set until 2001. Consequently, the analyses performed in this study likely underestimate the amount of training required

for less capable personnel to fill the current job of Predator operator. As more training requirements are identified, the cost of these less capable personnel will increase accordingly, thereby decreasing the observed difference between them and experienced, rated Air Force aviators in terms of overall cost effectiveness.

While the observed difference in cost effectiveness between the optimal and the other feasible solutions was an order of magnitude, it would be prudent to perform a sensitivity analysis to predict the effects of model parameter changes on the choice of an optimal solution. As an example, cutting the training time limit in half (i.e., tightening the constraint) makes Predator selectees the optimal choice, while doubling the limit (i.e., relaxing the constraint) makes civil private pilots the optimal choice. Since the choice of many of these parameters is arbitrary for the purpose of demonstrating the analysis, it is beyond the scope of this discussion to consider an exhaustive sensitivity analysis. Nonetheless, these issues are quite important in practice because we often have incomplete knowledge of the data related to the problem (Rardin, 1998).

These types of issues will inevitably cause analysis problems for other systems, particularly if human-related data are obtained from simulation-based experiments employing parameterized human performance models rather than human-in-the-loop experiments. Nevertheless, despite these problems, a relatively straightforward example of the type of HSI related research and analysis that needs to be conducted to improve personnel utilization is provided in this study. A demonstration of the need for feedback of personnel factors to designers in the early stages of the acquisition process is also provided. To summarize, this study illustrates the importance of integrating human factors research and experimentation with system analysis so that Air Force personnel managers may improve total service performance through optimized selection, recruiting, and allocation policies.

F. APPENDIX

1. Basic Maneuvering Task Regression Model

Our initial logistic regression model relates the proportion proficient to the two predictor variables, `Time` and `Group`, as well as their interaction, `Time:Group`. We fit the general linear model using `glm` as follows:

```
> bmtask.glm.all <- glm(y ~ Time + Group + Time:Group, weight = Weight,
+ family = binomial(logit), data = bmtask.df)
```

The summary of the resulting fit:

```
> summary(bmtask.glm.all)

Call:  glm(formula  =  y  ~  Time  +  Group  +  Time:Group,  family  =
binomial(logit), data = bmtask.df, weights = Weight)
Deviance Residuals:
       Min         1Q       Median        3Q      Max
 -1.455336 -0.3584625 -0.008955164 0.3422195 1.392584

Coefficients:
                             Value Std. Error    t value
            (Intercept) -3.79655561 0.58062723 -6.5387144
                   Time  0.07078031 0.01099094  6.4398757
        GroupCivil_inst -1.62448130 0.98019136 -1.6573104
        GroupCivil_priv -1.85765153 1.18203188 -1.5715748
     GroupPred_selectee -3.71232676 1.11643460 -3.3251628
           GroupT1_grad -1.43579321 0.92641172 -1.5498435
          GroupT38_grad -0.76651247 0.88977081 -0.8614718
    TimeGroupCivil_inst  0.05546095 0.02103663  2.6363990
    TimeGroupCivil_priv  0.04120973 0.02320528  1.7758772
TimeGroupPred_selectee  0.09976107 0.02346270  4.2519012
       TimeGroupT1_grad  0.03177690 0.01787354  1.7778734
      TimeGroupT38_grad  0.02532952 0.01745794  1.4508885

(Dispersion Parameter for Binomial family taken to be 1 )

    Null Deviance: 557.0008 on 90 degrees of freedom

Residual Deviance: 25.35335 on 79 degrees of freedom

Number of Fisher Scoring Iterations: 5
```

The test that all slopes are zero, G = 531.6475, DF = 11, and P-value < 0.001, indicates the model is adequate. The partial *t*-tests show that `Time` is important even after adjusting for `Group` and the interaction terms, but they provide less information on `Groups` or the interaction although these too seem to be important.

We next examine the bivariate relationships between the proportion proficient and each of the predictors by fitting a "null" model and then adding each of the terms, one at a time:

```
> bmtask.glm.null <- glm(y ~ 1, weight = Weight, family =
+ binomial(logit), data = bmtask.df)

> add1(bmtask.glm.null, ~ . + Time + Group)

Single term additions

Model:
y ~ 1
       Df Sum of Sq      RSS       Cp
<none>              475.8387 486.4129
  Time  1  309.1494 166.6893 187.8377
 Group  5    0.2506 475.5882 539.0333
```

Using the `Cp` statistic to compare the models, `Time` is clearly the best single variable to use in a linear model. However, to examine the contribution of `Group` and the interaction term to the full model, we produce an analysis of deviance for the sequential addition of each variable by using the `anova` function and specifying the chi-square test to evaluate for differences between models:

```
> anova(bmtask.glm.all, test = "Chisq")

Analysis of Deviance Table

Binomial model

Response: y

Terms added sequentially (first to last)
           Df Deviance Resid. Df Resid. Dev      Pr(Chi)
      NULL                    90   557.0008
      Time  1 468.9139        89    88.0869 0.0000000000
     Group  5  40.7733        84    47.3136 0.0000001043
Time:Group  5  21.9603        79    25.3534 0.0005327781
```

Here we see that `Group` is important after adjusting for `Time`, and the interaction term, `Time:Group`, is important after adjusting for both `Time` and `Group`.

These statistical conclusions are subsequently verified by looking at graphical displays of the fitted values and residuals (Figure V-26). There does not appear to be significant systemic curvature or other patterns in the plot of residuals against fitted values (Figure V-26A). There also does not appear to be any large residuals suggesting outlying observations that might skew the analysis. The plot of the absolute residuals

against predicted values suggests the assumed variance function is adequate (Figure V-26B). The normal quantile plot indicates the left tail of the distribution is light (Figure V-26D), suggesting some problems with the model fit, but this must be tempered in the non-linear context.

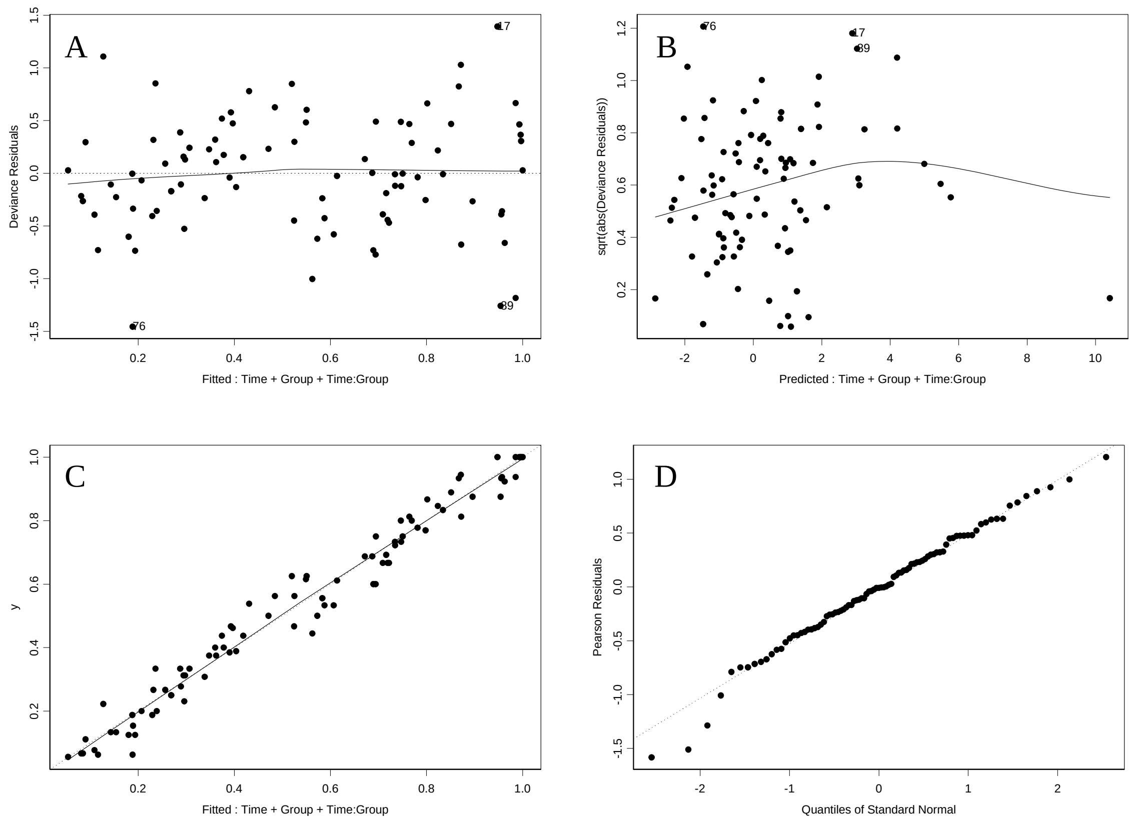


Figure V-26. Plots of the generalized linear model of `Proficient` (*y*) predicted by `Time`, `Group`, and `Time:Group`.

Tables V-6 thru V-8 contain the output from S-Plus, including the leverages calculated using the `lm` function:

```
> h.i. <-lm.influence(bmtask.glm.all)$hat
```

The maximum standardized deviance residual is 1.47574 and the maximum standardized pearson residual is 1.33038, suggesting there are no outliers in the dataset. Likewise, the maximum observed leverage is 0.23674, which is below the calculated h_{max} = 0.26374 threshold where one becomes concerned about "excessive" leverage. Likewise,

looking at Cook's D, D_{max} for the entire dataset is 0.05545, which is well below the threshold for concern for an unduly influential observation.

Table V-6. Residuals for the basic maneuvering task data.

Observation	Group	Time	Observed Probability	Estimated Probability	Deviance Residuals	Pearson Residuals	h_{ii}	Cook's Distances
1	Pred_selectee	27.25	0.05556	0.05408	0.02991	0.03004	0.14973	0.00001
2	Pred_selectee	30.50	0.11111	0.09051	0.32427	0.33471	0.17210	0.00194
3	Pred_selectee	32.75	0.22222	0.12745	1.22282	1.33038	0.17875	0.03210
4	Pred_selectee	38.75	0.27778	0.28897	-0.11363	-0.11319	0.14412	0.00018
5	Pred_selectee	39.25	0.33333	0.30680	0.26098	0.26298	0.13847	0.00093
6	Pred_selectee	41.75	0.38889	0.40401	-0.13888	-0.13859	0.11012	0.00020
7	Pred_selectee	45.50	0.44444	0.56236	-1.04916	-1.05464	0.08565	0.00868
8	Pred_selectee	45.75	0.50000	0.57282	-0.64956	-0.65306	0.08527	0.00331
9	Pred_selectee	46.00	0.55556	0.58322	-0.24816	-0.24887	0.08504	0.00048
10	Pred_selectee	46.75	0.61111	0.61395	-0.02582	-0.02584	0.08530	0.00001
11	Pred_selectee	49.25	0.66667	0.70895	-0.40944	-0.41488	0.09413	0.00149
12	Pred_selectee	50.00	0.72222	0.73462	-0.12486	-0.12546	0.09825	0.00014
13	Pred_selectee	51.50	0.77778	0.78143	-0.03953	-0.03961	0.10684	0.00002
14	Pred_selectee	53.50	0.83333	0.83412	-0.00953	-0.00953	0.11648	0.00000
15	Pred_selectee	54.25	0.88889	0.85107	0.49922	0.48022	0.11900	0.00260
16	Pred_selectee	55.25	0.94444	0.87142	1.09709	0.98732	0.12122	0.01121
17	Pred_selectee	61.00	1.00000	0.94756	1.47574	1.05772	0.10952	0.01147
18	T38_grad	22.75	0.06667	0.08498	-0.28920	-0.27915	0.17034	0.00133
19	T38_grad	29.75	0.13333	0.15397	-0.24855	-0.24380	0.17474	0.00105
20	T38_grad	33.50	0.20000	0.20695	-0.07300	-0.07269	0.16375	0.00009
21	T38_grad	35.00	0.26667	0.23161	0.34479	0.35053	0.15696	0.00191
22	T38_grad	35.25	0.33333	0.23592	0.92913	0.96710	0.15572	0.01438
23	T38_grad	41.50	0.40000	0.36020	0.33995	0.34246	0.12073	0.00134
24	T38_grad	48.50	0.46667	0.52454	-0.47263	-0.47299	0.09953	0.00206
25	T38_grad	52.00	0.53333	0.60698	-0.61166	-0.61717	0.10451	0.00370
26	T38_grad	56.00	0.60000	0.69404	-0.82255	-0.84259	0.12007	0.00807
27	T38_grad	57.25	0.66667	0.71894	-0.47353	-0.48173	0.12596	0.00279
28	T38_grad	58.75	0.73333	0.74713	-0.13133	-0.13210	0.13305	0.00022
29	T38_grad	60.00	0.80000	0.76915	0.31063	0.30552	0.13870	0.00125
30	T38_grad	62.00	0.86667	0.80151	0.71859	0.68496	0.14670	0.00672
31	T38_grad	67.00	0.93333	0.86718	0.89843	0.82260	0.15791	0.01057
32	T38_grad	107.50	1.00000	0.99689	0.31079	0.21993	0.03133	0.00013
33	T1_grad	31.25	0.06250	0.11635	-0.79401	-0.73143	0.15642	0.00827
34	T1_grad	36.25	0.12500	0.18025	-0.65228	-0.62316	0.14879	0.00566
35	T1_grad	36.75	0.18750	0.18795	-0.00502	-0.00501	0.14717	0.00000
36	T1_grad	41.25	0.25000	0.26858	-0.18088	-0.17952	0.12796	0.00039
37	T1_grad	42.50	0.31250	0.29449	0.16769	0.16866	0.12202	0.00033
38	T1_grad	45.50	0.37500	0.36216	0.11295	0.11323	0.10952	0.00013
39	T1_grad	46.00	0.43750	0.37409	0.54959	0.55498	0.10789	0.00310
40	T1_grad	52.00	0.56250	0.52514	0.31709	0.31654	0.10600	0.00099
41	T1_grad	53.00	0.62500	0.55062	0.63819	0.63371	0.10922	0.00410

Table V-7. Residuals for the basic maneuvering task data (continued).

Observation	Group	Time	Observed Probability	Estimated Probability	Deviance Residuals	Pearson Residuals	h_{ii}	Cook's Distances
42	T1_grad	58.00	0.68750	0.67172	0.14526	0.14465	0.13633	0.00028
43	T1_grad	61.75	0.75000	0.75036	-0.00368	-0.00368	0.16019	0.00000
44	T1_grad	69.75	0.81250	0.87225	-0.74902	-0.79306	0.18491	0.01189
45	T1_grad	72.00	0.87500	0.89583	-0.29359	-0.30178	0.18280	0.00170
46	T1_grad	81.25	0.93750	0.95691	-0.38850	-0.41363	0.14548	0.00243
47	T1_grad	99.75	1.00000	0.99329	0.47748	0.33820	0.05530	0.00056
48	Civil_inst	23.75	0.06667	0.08145	-0.23585	-0.22899	0.16474	0.00086
49	Civil_inst	28.75	0.13333	0.14288	-0.11710	-0.11599	0.17020	0.00023
50	Civil_inst	33.75	0.20000	0.23860	-0.38780	-0.38012	0.14878	0.00210
51	Civil_inst	34.50	0.26667	0.25622	0.09976	0.10020	0.14373	0.00014
52	Civil_inst	35.75	0.33333	0.28743	0.41621	0.42233	0.13481	0.00232
53	Civil_inst	39.00	0.40000	0.37810	0.18504	0.18571	0.11271	0.00037
54	Civil_inst	39.50	0.46667	0.39305	0.61325	0.61877	0.10995	0.00394
55	Civil_inst	45.75	0.53333	0.58771	-0.44979	-0.45227	0.10506	0.00200
56	Civil_inst	49.25	0.60000	0.68919	-0.78031	-0.79785	0.12488	0.00757
57	Civil_inst	50.50	0.66667	0.72195	-0.50395	-0.51333	0.13349	0.00338
58	Civil_inst	51.00	0.73333	0.73444	-0.01047	-0.01048	0.13692	0.00000
59	Civil_inst	51.50	0.80000	0.74657	0.52653	0.51311	0.14029	0.00358
60	Civil_inst	54.00	0.86667	0.80155	0.72174	0.68798	0.15507	0.00724
61	Civil_inst	67.25	0.93333	0.95558	-0.42034	-0.45021	0.13663	0.00267
62	Civil_inst	76.25	1.00000	0.98530	0.69603	0.49399	0.08272	0.00183
63	Civil_priv	31.75	0.07692	0.10924	-0.44172	-0.42069	0.21148	0.00396
64	Civil_priv	37.50	0.15385	0.18930	-0.37217	-0.36263	0.19059	0.00258
65	Civil_priv	42.75	0.23077	0.29596	-0.57110	-0.55746	0.14692	0.00446
66	Civil_priv	44.50	0.30769	0.33835	-0.25286	-0.25089	0.13269	0.00080
67	Civil_priv	46.50	0.38462	0.39015	-0.04364	-0.04360	0.12015	0.00002
68	Civil_priv	46.75	0.46154	0.39683	0.50412	0.50804	0.11897	0.00290
69	Civil_priv	48.00	0.53846	0.43078	0.82820	0.83330	0.11467	0.00750
70	Civil_priv	52.25	0.61538	0.54916	0.51530	0.51224	0.12247	0.00305
71	Civil_priv	58.75	0.69231	0.71610	-0.20919	-0.21089	0.18609	0.00085
72	Civil_priv	62.75	0.76923	0.79789	-0.28770	-0.29253	0.22614	0.00208
73	Civil_priv	64.25	0.84615	0.82363	0.24816	0.24383	0.23674	0.00154
74	Civil_priv	79.50	0.92308	0.96264	-0.73532	-0.83646	0.19166	0.01382
75	Civil_priv	143.50	1.00000	0.99997	0.02792	0.01974	0.00145	0.00000
76	Cadet	33.00	0.06250	0.18834	-1.57956	-1.39735	0.15111	0.02896
77	Cadet	33.50	0.12500	0.19381	-0.79597	-0.75489	0.14908	0.00832
78	Cadet	36.50	0.18750	0.22916	-0.43645	-0.42645	0.13568	0.00238
79	Cadet	39.50	0.25000	0.26880	-0.18231	-0.18092	0.12126	0.00038
80	Cadet	41.50	0.31250	0.29751	0.13850	0.13916	0.11192	0.00020
81	Cadet	44.75	0.37500	0.34771	0.24011	0.24150	0.09891	0.00053
82	Cadet	49.00	0.43750	0.41865	0.15986	0.16018	0.08950	0.00021
83	Cadet	52.00	0.50000	0.47104	0.24310	0.24331	0.09008	0.00049
84	Cadet	52.75	0.56250	0.48428	0.65701	0.65674	0.09126	0.00361
85	Cadet	54.75	0.62500	0.51966	0.89314	0.88727	0.09645	0.00700
86	Cadet	64.75	0.68750	0.68707	0.00401	0.00401	0.15550	0.00000

Table V-8. Residuals for the basic maneuvering task data (continued).

Observation	Group	Time	Observed Probability	Estimated Probability	Deviance Residuals	Pearson Residuals	h_{ii}	Cook's Distances
87	Cadet	65.25	0.75000	0.69463	0.53440	0.52442	0.15914	0.00434
88	Cadet	70.25	0.81250	0.76418	0.52029	0.50691	0.19341	0.00513
89	Cadet	96.50	0.87500	0.95408	-1.39855	-1.67982	0.19081	0.05545
90	Cadet	113.00	0.93750	0.98525	-1.25523	-1.68233	0.11308	0.03007
91	Cadet	131.00	1.00000	0.99583	0.37573	0.26596	0.05282	0.00033

We next examine several alternative models with different link functions and scaling of predictors to determine if the model fit can be improved. We fit models using the complementary log-log (`cloglog`) and `probit` link functions:

```
> bmtask.glm.cloglog <- glm(y ~ Time + Group + Time:Group, weight =
+ Weight, family = binomial(cloglog), data = bmtask.df)

> bmtask.glm.probit <- glm(y ~ Time + Group + Time:Group, weight =
+ Weight, family = binomial(probit), data = bmtask.df)
```

We also evaluate rescaling the predictor variable, `Time`, using a log transformation:

```
> bmtask.glm.logtime <- glm(y ~ log(Time) + Group + log(Time):Group,
+ weight = Weight, family = binomial(logit), data = bmtask.df)
```

We then use the `anova` function to compare these models and look at the graphical displays of the fitted values and residuals:

```
> anova(bmtask.glm.all, bmtask.glm.logtime, bmtask.glm.cloglog,
+ bmtask.glm.probit)

Analysis of Deviance Table

Response: y

                               Terms Resid. Df Resid. Dev
1           Time + Group + Time:Group        79   25.35335
2 log(Time) + Group + log(Time):Group        79   25.40528
3           Time + Group + Time:Group        79   39.11041
4           Time + Group + Time:Group        79   26.70424
```

The residual deviances suggest that the model using the `logit` link function (#1) is the best, although the models using the rescaled predictor, log(`Time`), (#2) and the `probit` link function (#4) are comparable. Examining the graphical displays for the model with the rescaled predictor (Figure V-27), we see from the normal quantile plot (Figure V-27D) that the problem of the light tailed distribution is not improved. Also, introduction

of the rescaled parameter has led to some systemic curvature in the plot of residuals against fitted values (Figure V-27A). Overall, our best fit appears to be obtained with the initial model.

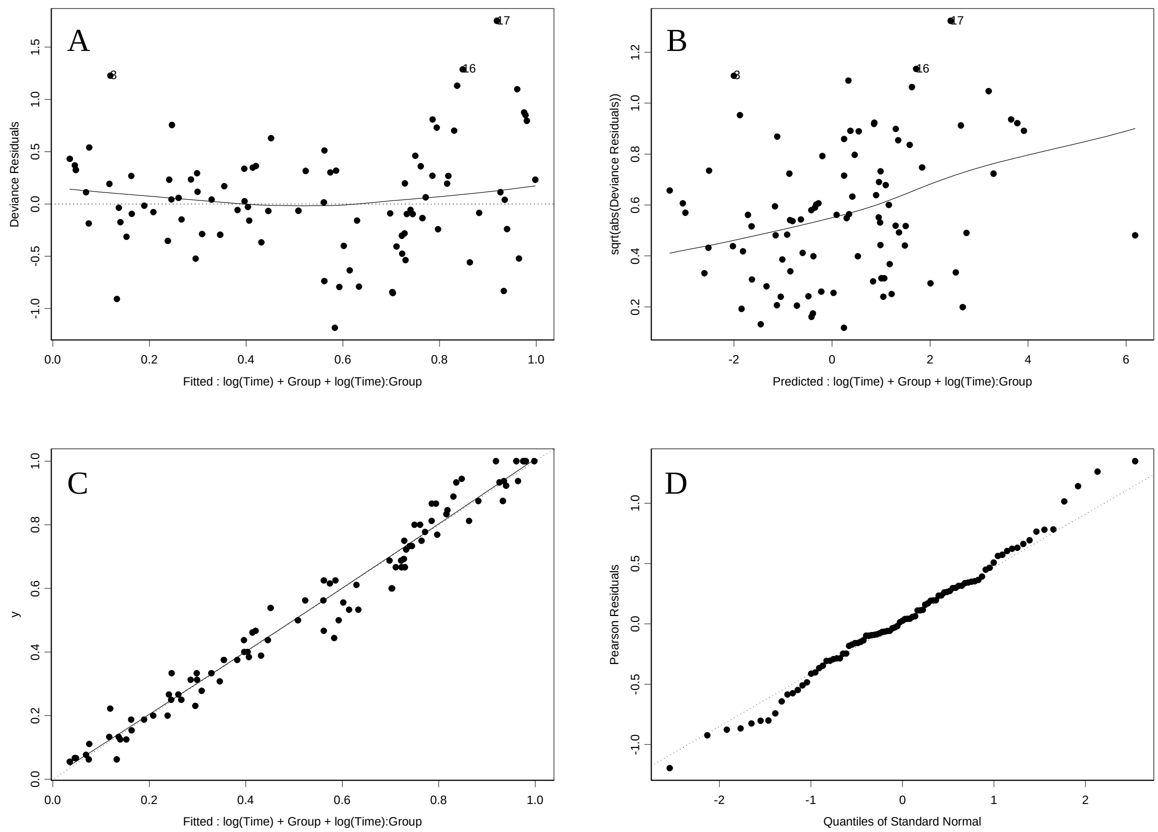


Figure V-27. Plots of the generalized linear model of `Proficient` (*y*) predicted by log(`Time`), `Group`, and log(`Time`)`:Group`.

So far, we have examined only linear relationships between the predictors and the proportion proficient. We now assess the validity of the linear assumption by fitting an additive model with relationships estimated by smoothing operations, and then comparing the linear fit. We first use the `gam` function to fit an additive model, indicating `Time` as an argument to the `s` function, to estimate a “smoothed” relationship as follows:

```
> bmtask.gam.all <- gam(y ~ s(Time) + Group + s(Time):Group, weight =
+ Weight, family = binomial(logit), data = bmtask.df)
```

A summary of the fit is:

```
> summary(landing.gam.all)

Call:  gam(formula  =  y  ~  s(Time)  +  Group  +  s(Time):Group,  family  =
binomial(logit), data = bmtask.df, weights = Weight)
Deviance Residuals:

  Min         1Q      Median         3Q      Max
 -1.146995 -0.2885017 -0.02004951 0.3320611 1.431841

(Dispersion Parameter for Binomial family taken to be 1 )

    Null Deviance: 557.0008 on 90 degrees of freedom

Residual Deviance: 21.26504 on 76.06075 degrees of freedom

Number of Local Scoring Iterations: 6

DF for Terms and Chi-squares for Nonparametric Effects

              Df Npar Df Npar Chisq   P(Chi)
  (Intercept)  1
      s(Time)  1     2.9   4.544914 0.201098
        Group  5
s(Time):Group  5
```

Since the non-parametric tests do not inform us about the contribution of `Group` and the interaction term in the presence of a smooth of `Time`, we fit two additional models that build on a base model: one with the `Group` variable and one with a smooth of the `Time:Group` variable.

```
> bmtask.gam.time <- gam(y ~ s(Time), weight = Weight, family =
+ binomial(logit), data = bmtask.df)

> bmtask.gam.time.group <- gam(y ~ s(Time) + Group, weight = Weight,
+ family = binomial(logit), data = bmtask.df)
```

We then produce the following analysis of deviance table:

```
> anova(bmtask.gam.time, bmtask.gam.time.group, bmtask.gam.all, test =
+ "Chisq")

Analysis of Deviance Table

Response: y

                            Terms Resid. Df Resid. Dev
1                         s(Time)  86.04682   72.07984
2                 s(Time) + Group  81.04223   33.98506
3 s(Time) + Group + s(Time):Group  76.06075   21.26504

Test                 Df Deviance    Pr(Chi)
+Group         5.004587 38.09478 0.00000036
+s(Time):Group 4.981476 12.72001 0.0258325
```

The indication is that `Group` is important in the model even with `Time` included, and the interaction term, `Time:Group`, is important even in the presence of both `Time` and `Group`. Figure V-28 shows the graphical displays for the plots of the partial residuals (Figure V-28A) and the pointwise confidence intervals for the model that includes the `Time` and `Group` variables and interaction term (Figure V-28B). The plots suggest a possible piecewise linear relationship for `Time`.

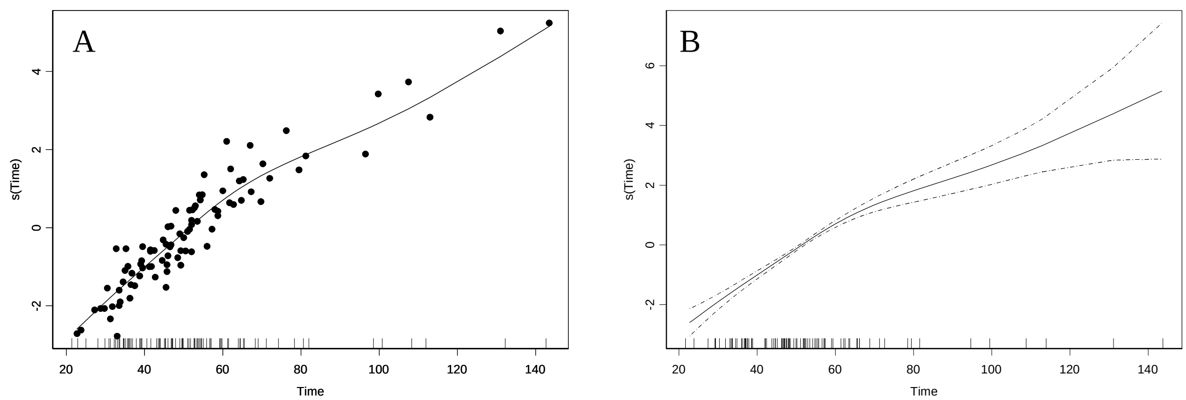


Figure V-28. The partial fits for the generalized additive logistic regression model of `Proficient` (y) with `Time`, `Group`, and `Time:Group` as predictors.

We next use the `anova` function to compare the linear fit `bmtask.glm.all` with the additive fit `bmtask.gam.all` to investigate whether it may be worthwhile proceeding to develop a more complex model:

```
> anova(bmtask.glm.all, bmtask.gam.all, test = "Chisq")

Analysis of Deviance Table

Response: y

                            Terms Resid. Df Resid. Dev
1        Time + Group + Time:Group  79.00000   25.35335
2 s(Time) + Group + s(Time):Group  76.06075   21.26504

   Test       Df Deviance   Pr(Chi)

1 vs. 2 2.939245 4.088307 0.2438847
```

We see that the linear fit is more parsimonious. The effective degrees of freedom are 79 with the linear model and approximately 76 in the additive model with smooths. The

residual deviance in the linear fit is not significantly higher than the residual deviance in the additive fit. In addition, with the linear fit, we can produce an analytical expression for the model, which cannot be done for an additive model with smooth fits. Given these considerations, we decide to use the linear fit to develop our subsequent model.

2. Landing Task Regression Model

Our initial logistic regression model relates the proportion proficient to the two predictor variables, `Group` and `Trials`, as well as their interaction, `Trials:Group`. We fit the general linear model using `glm` as follows:

```
> landing.glm.all <- glm(y ~ Trials + Group + Trials:Group, weight =
+ Weight, family = binomial(logit), data = landing.task.df)
```

The summary of the resulting fit:

```
> summary(landing.glm.all)

Call: glm(formula = y ~ Trials + Group + Trials:Group, weight = Weight,
+ family = binomial(logit), data = landing.task.df)
Deviance Residuals:
      Min         1Q     Median         3Q      Max
 -2.10377 -0.3117453 0.01403171 0.3196447 1.365283

Coefficients:
                                Value  Std. Error    t value
             (Intercept) -3.597614482 0.536623518 -6.7041685
                  Trials  0.032536538 0.004829464  6.7370912
         GroupCivil_inst  0.835137384 0.775831095  1.0764423
         GroupCivil_priv  1.827838889 0.629637465  2.9030021
      GroupPred_selectee -0.471719851 0.745265420 -0.6329555
            GroupT1_grad  1.517845511 0.616748497  2.4610445
           GroupT38_grad  0.778441653 0.685203581  1.1360735
   TrialsGroupCivil_inst  0.031685596 0.013298311  2.3826782
   TrialsGroupCivil_priv -0.002508798 0.006838017 -0.3668896
TrialsGroupPred_selectee  0.053102199 0.011633863  4.5644511
      TrialsGroupT1_grad -0.005418159 0.006020155 -0.9000031
     TrialsGroupT38_grad  0.019976847 0.008910008  2.2420683

(Dispersion Parameter for Binomial family taken to be 1 )

    Null Deviance: 551.8535 on 88 degrees of freedom

Residual Deviance: 30.27233 on 77 degrees of freedom

Number of Fisher Scoring Iterations: 5
```

The test that all slopes are zero, G = 521.5812, DF = 11, and P-value < 0.001, indicates the model is adequate. The partial *t*-tests show that `Trials` is important even after adjusting for `Group` and the interaction terms, but they provide little information on `Groups` or the interaction.

We next examine the bivariate relationships between the proportion proficient and each of the predictors by fitting a "null" model and then adding each of the terms, one at a time. Our null model has a single intercept term and is specified with the formula: `y ~ 1`:

```
> landing.glm.null <- glm(y ~ 1, weight = Weight, family =
+ binomial(logit), data = landing.task.df)

> add1(landing.glm.null, ~ . + Trials + Group)

Single term additions

Model: y ~ 1
       Df Sum of Sq      RSS       Cp
<none>              470.2589 480.9466
Trials  1  283.7513 186.5076 207.8830
 Group  5    1.8372 468.4217 532.5479
```

The `Cp` statistic is used to compare the models. A small `Cp` value corresponds to a better model in the sense of a smaller residual deviance penalized by the numbers of parameters that are estimated in fitting the model. From the above analysis, `Trials` is clearly the best single variable to use in a linear model. However, to examine the contribution of `Group` and the interaction term to the full model, we produce an analysis of deviance for the sequential addition of each variable by using the `anova` function and specifying the chi-square test to evaluate for differences between models:

```
> anova(landing.glm.all, test = "Chisq")

Analysis of Deviance Table

Binomial model

Response: y

Terms added sequentially (first to last)
             Df Deviance Resid. Df Resid. Dev      Pr(Chi)
        NULL                    88   551.8535
      Trials  1 341.8164        87   210.0372 0.000000e+000
       Group  5 132.5372        82    77.5000 0.000000e+000
Trials:Group  5  47.2276        77    30.2723 5.105621e-009
```

Here we see that `Group` is important after adjusting for `Trials`, and the interaction term, `Trials:Group`, is important after adjusting for both `Trials` and `Group`.

These statistical conclusions are subsequently verified by looking at graphical displays (Figure V-29) of the fitted values and residuals. The plots indicate there may be some problems with the model fit. The slight systematic curvature in the plot of deviance residuals versus the estimated proportion proficient (Figure V-29A) may be indicative of problems in the choice of link, the wrong scale for a predictor, or an omission of a quadratic term in the predictor. There do not appear to be large residuals suggesting outlying observations that might skew the analysis. The plot of the absolute residuals against predicted values (Figure V-29B) suggests the assumed variance function is adequate. The normal quantile plot (Figure V-29D) indicates one possible extreme observation, but this must be tempered in the non-linear context.

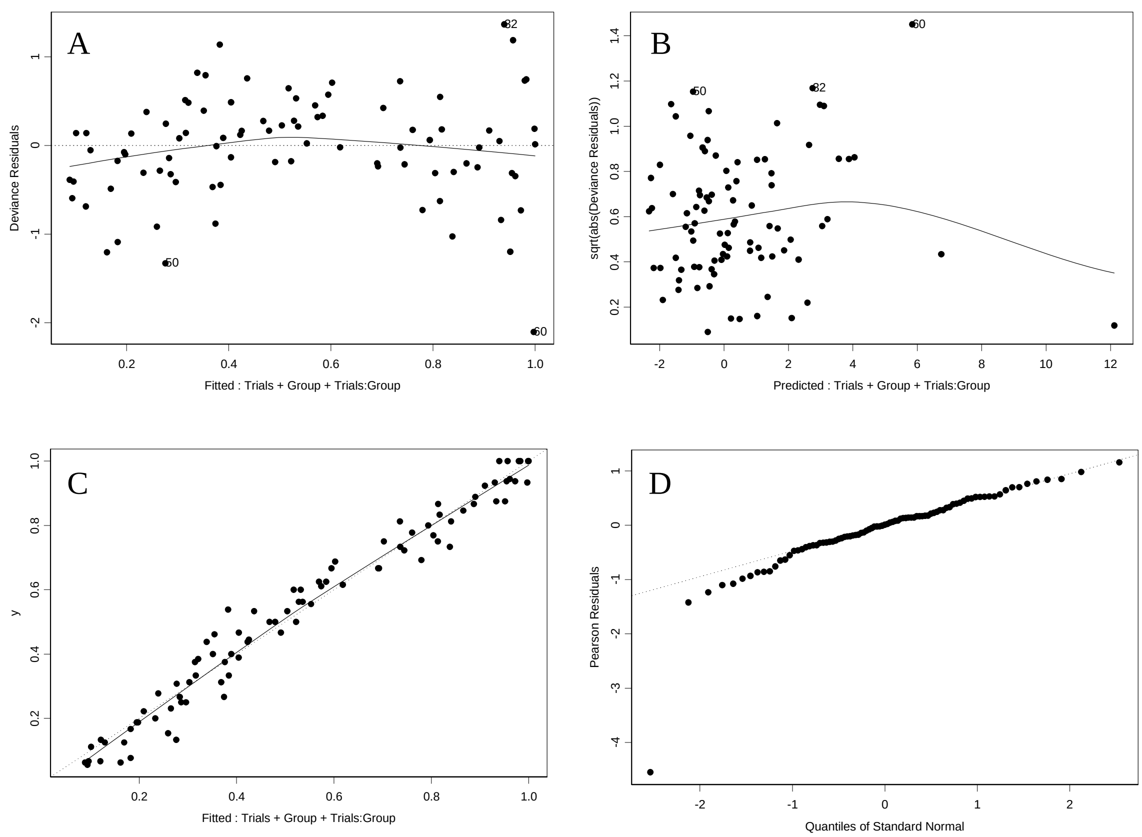


Figure V-29. Plots of the generalized linear model of `Proficient` (*y*) predicted by `Trials`, `Group`, and `Trials:Group`.

To help interpret this observation, it is useful to compare leverages and residuals since observations with large leverages and large residuals are likely to be influential. Tables V-9 and V-10 contain the output from S-Plus, including the leverages calculated using the `lm` function:

```
> h.i. <-lm.influence(landing.glm.all)$hat
```

The apparently extreme observation is observation 60, corresponding to y = 0.93333, which is not an outlier observation for `GroupCivil_inst`. Additionally, the leverage for this observation, h_{60} = 0.05618, is well below h_{max} = 0.26966 and so is not considered "excessive." There is only one observation with excessive leverage: h_{49} = 0.28119. However, observation 49 corresponds to y = 0.75000, and hence is not an outlier observation for `GroupT1_grad`. Cook's D (D_i) is another statistic used to assess the influence of an individual observation. We usually consider observations for which D_i > 1 to be influential. In this case, D_{max} for the entire dataset is 0.11124, which is well below the threshold for concern.

Table V-9. Residuals for the landing task data.

Observation	Group	Trials	Observed Probability	Estimated Probability	Deviance Residuals	Pearson Residuals	h_{ii}	Cook's Distances
1	Pred_selectee	21	0.05556	0.09356	-0.64342	-0.59875	0.14510	0.00507
2	Pred_selectee	22	0.11111	0.10108	0.15062	0.15277	0.14607	0.00033
3	Pred_selectee	30	0.16667	0.18240	-0.18830	-0.18613	0.13801	0.00046
4	Pred_selectee	32	0.22222	0.20934	0.14306	0.14411	0.13162	0.00026
5	Pred_selectee	34	0.27778	0.23910	0.40403	0.41105	0.12392	0.00199
6	Pred_selectee	42	0.33333	0.38402	-0.46827	-0.46390	0.09135	0.00180
7	Pred_selectee	43	0.38889	0.40447	-0.14139	-0.14108	0.08851	0.00016
8	Pred_selectee	44	0.44444	0.42526	0.17190	0.17222	0.08622	0.00023
9	Pred_selectee	46	0.50000	0.46756	0.28786	0.28816	0.08352	0.00063
10	Pred_selectee	50	0.55556	0.55295	0.02329	0.02329	0.08635	0.00000
11	Pred_selectee	51	0.61111	0.57401	0.33484	0.33342	0.08873	0.00090
12	Pred_selectee	57	0.66667	0.69255	-0.25074	-0.25261	0.11259	0.00067
13	Pred_selectee	60	0.72222	0.74440	-0.22881	-0.23091	0.12675	0.00064
14	Pred_selectee	61	0.77778	0.76036	0.18740	0.18580	0.13121	0.00043
15	Pred_selectee	65	0.83333	0.81715	0.19440	0.19210	0.14581	0.00052
16	Pred_selectee	72	0.88889	0.89058	-0.02488	-0.02493	0.15385	0.00001
17	Pred_selectee	85	0.94444	0.96121	-0.36961	-0.39275	0.12017	0.00176
18	Pred_selectee	189	1.00000	0.99999	0.01403	0.00992	0.00022	0.00000
19	T38_grad	11	0.06667	0.09608	-0.44443	-0.42210	0.16139	0.00286
20	T38_grad	16	0.13333	0.12143	0.15220	0.15428	0.16348	0.00039
21	T38_grad	31	0.20000	0.23303	-0.33294	-0.32714	0.14425	0.00150
22	T38_grad	36	0.26667	0.28319	-0.15334	-0.15241	0.13183	0.00029
23	T38_grad	39	0.33333	0.31623	0.15149	0.15220	0.12433	0.00027
24	T38_grad	42	0.40000	0.35124	0.41715	0.42116	0.11752	0.00197
25	T38_grad	53	0.46667	0.49101	-0.19964	-0.19954	0.10694	0.00040
26	T38_grad	54	0.53333	0.50414	0.23947	0.23936	0.10749	0.00058
27	T38_grad	55	0.60000	0.51726	0.68201	0.67913	0.10832	0.00467
28	T38_grad	69	0.66667	0.69088	-0.21753	-0.21903	0.14168	0.00066
29	T38_grad	85	0.73333	0.83814	-1.13473	-1.21736	0.18037	0.02718
30	T38_grad	93	0.86667	0.88741	-0.27406	-0.28104	0.18225	0.00147
31	T38_grad	103	0.93333	0.93020	0.05260	0.05223	0.16841	0.00005
32	T38_grad	106	1.00000	0.93976	1.49121	1.07104	0.16176	0.01845
33	T1_grad	16	0.06250	0.16167	-1.30140	-1.16317	0.14191	0.01865
34	T1_grad	18	0.12500	0.16915	-0.52828	-0.50820	0.14062	0.00352
35	T1_grad	25	0.18750	0.19753	-0.10904	-0.10833	0.13486	0.00015
36	T1_grad	43	0.25000	0.28625	-0.34546	-0.34081	0.11416	0.00125
37	T1_grad	46	0.31250	0.30315	0.08600	0.08624	0.11051	0.00008
38	T1_grad	48	0.37500	0.31473	0.54126	0.54966	0.10814	0.00305
39	T1_grad	52	0.43750	0.33858	0.86614	0.88313	0.10368	0.00752
40	T1_grad	80	0.50000	0.52241	-0.18851	-0.18860	0.09458	0.00031
41	T1_grad	82	0.56250	0.53592	0.22453	0.22419	0.09587	0.00044
42	T1_grad	87	0.62500	0.56943	0.47627	0.47324	0.10022	0.00208
43	T1_grad	92	0.68750	0.60232	0.74755	0.73631	0.10595	0.00535
44	T1_grad	131	0.75000	0.81347	-0.68458	-0.71146	0.16070	0.00808
45	T1_grad	138	0.81250	0.84058	-0.32854	-0.33594	0.16567	0.00187
46	T1_grad	174	0.87500	0.93332	-0.91390	-1.01640	0.15358	0.01562

Table V-10. Residuals for the landing task data (continued).

Observation	Group	Trials	Observed Probability	Estimated Probability	Deviance Residuals	Pearson Residuals	h_{ii}	Cook's Distances
47	T1_grad	189	0.93750	0.95459	-0.33498	-0.35330	0.13607	0.00164
48	T1_grad	191	1.00000	0.95689	1.27574	0.91212	0.13349	0.01068
49	Civil_inst	12	0.06667	0.12007	-0.81109	-0.75042	0.28119	0.01836
50	Civil_inst	28	0.13333	0.27603	-1.47913	-1.37569	0.19239	0.03757
51	Civil_inst	35	0.26667	0.37409	-0.94933	-0.92618	0.13822	0.01147
52	Civil_inst	36	0.40000	0.38925	0.09154	0.09169	0.13207	0.00011
53	Civil_inst	37	0.46667	0.40462	0.52038	0.52392	0.12656	0.00331
54	Civil_inst	39	0.53333	0.43590	0.80594	0.81015	0.11774	0.00730
55	Civil_inst	45	0.60000	0.53184	0.56397	0.56143	0.11198	0.00331
56	Civil_inst	49	0.66667	0.59494	0.61233	0.60534	0.12591	0.00440
57	Civil_inst	59	0.73333	0.73626	-0.02876	-0.02880	0.19937	0.00002
58	Civil_inst	64	0.80000	0.79376	0.06867	0.06841	0.23763	0.00012
59	Civil_inst	66	0.86667	0.81400	0.63109	0.60555	0.25038	0.01021
60	Civil_inst	134	0.93333	0.99711	-2.16548	-4.73563	0.05618	0.11124
61	Civil_inst	148	1.00000	0.99882	0.19100	0.13509	0.03037	0.00005
62	Civil_priv	9	0.07692	0.18250	-1.19471	-1.08060	0.16824	0.01968
63	Civil_priv	24	0.15385	0.25940	-0.99337	-0.94021	0.14724	0.01272
64	Civil_priv	25	0.23077	0.26521	-0.30882	-0.30429	0.14561	0.00131
65	Civil_priv	27	0.30769	0.27707	0.26349	0.26634	0.14234	0.00098
66	Civil_priv	34	0.38462	0.32108	0.51842	0.52646	0.13117	0.00349
67	Civil_priv	39	0.46154	0.35464	0.84556	0.86083	0.12411	0.00875
68	Civil_priv	43	0.53846	0.38259	1.21174	1.23228	0.11944	0.01716
69	Civil_priv	75	0.61538	0.61829	-0.02320	-0.02321	0.13302	0.00001
70	Civil_priv	101	0.69231	0.77955	-0.80665	-0.83989	0.18391	0.01325
71	Civil_priv	106	0.76923	0.80426	-0.34665	-0.35402	0.19124	0.00247
72	Civil_priv	121	0.84615	0.86571	-0.22717	-0.23163	0.20229	0.00113
73	Civil_priv	136	0.92308	0.91003	0.18774	0.18348	0.19670	0.00069
74	Civil_priv	188	1.00000	0.97968	0.77654	0.55193	0.11469	0.00329
75	Cadet	39	0.06250	0.08877	-0.42606	-0.40526	0.16858	0.00278
76	Cadet	52	0.12500	0.12946	-0.05861	-0.05832	0.16850	0.00006
77	Cadet	67	0.18750	0.19503	-0.08297	-0.08256	0.15265	0.00010
78	Cadet	84	0.25000	0.29639	-0.44077	-0.43355	0.12156	0.00217
79	Cadet	94	0.31250	0.36838	-0.49590	-0.49001	0.10577	0.00237
80	Cadet	95	0.37500	0.37598	-0.00856	-0.00856	0.10461	0.00000
81	Cadet	101	0.43750	0.42277	0.12555	0.12574	0.09985	0.00015
82	Cadet	108	0.50000	0.47910	0.17635	0.17643	0.09990	0.00029
83	Cadet	114	0.56250	0.52786	0.29391	0.29341	0.10513	0.00084
84	Cadet	121	0.62500	0.58402	0.35577	0.35387	0.11683	0.00138
85	Cadet	137	0.75000	0.70264	0.45915	0.45151	0.15739	0.00317
86	Cadet	142	0.81250	0.73547	0.79541	0.76707	0.17073	0.01010
87	Cadet	202	0.87500	0.95142	-1.32072	-1.56711	0.17676	0.04394
88	Cadet	220	0.93750	0.97236	-0.79011	-0.91748	0.14063	0.01148
89	Cadet	235	1.00000	0.98285	0.78913	0.56042	0.11109	0.00327

We next examine several alternative models with different link functions and scaling of predictors to determine if the model fit can be improved. We fit models using the complementary log-log (`cloglog`) and `probit` link functions:

```
> landing.glm.cloglog <- glm(y ~ Trials + Group + Trials:Group, weight
+ = Weight, family = binomial(cloglog), data = landing.task.df)

> landing.glm.probit <- glm(y ~ Trials + Group + Trials:Group, weight =
+ Weight, family = binomial(probit), data = landing.task.df)
```

We also evaluate rescaling the predictor variable, `Trials`, using a log transformation:

```
> landing.glm.logtrial <- glm(y ~ log(Trials) + Group
+ log(Trials):Group, weight = Weight, family = binomial(logit), data =
+ landing.task.df)
```

We then use the `anova` function to compare these models and look at the graphical displays of the fitted values and residuals:

```
> anova(landing.glm.all, landing.glm.logtrial, landing.glm.cloglog,
+ landing.glm.probit)

Analysis of Deviance Table

Response: y

                                    Terms Resid. Df Resid. Dev
1           Trials + Group + Trials:Group        77   30.27233
2 log(Trials) + Group + log(Trials):Group        77   31.99915
3           Trials + Group + Trials:Group        77   57.65410
4           Trials + Group + Trials:Group        77   34.50005
```

The residual deviances suggest that the model using the `logit` link function (#1) is the best, although the models using the rescaled predictor, `log(Trials)`, (#2) and the `probit` link function (#4) are comparable. Examining the graphical displays (Figure V-30) for the model with the rescaled predictor, we see that the systematic curvature in the residual plots is still present (Figure V-30A), but we no longer can detect an extreme observation on the normal quantile plot (Figure V-30D).

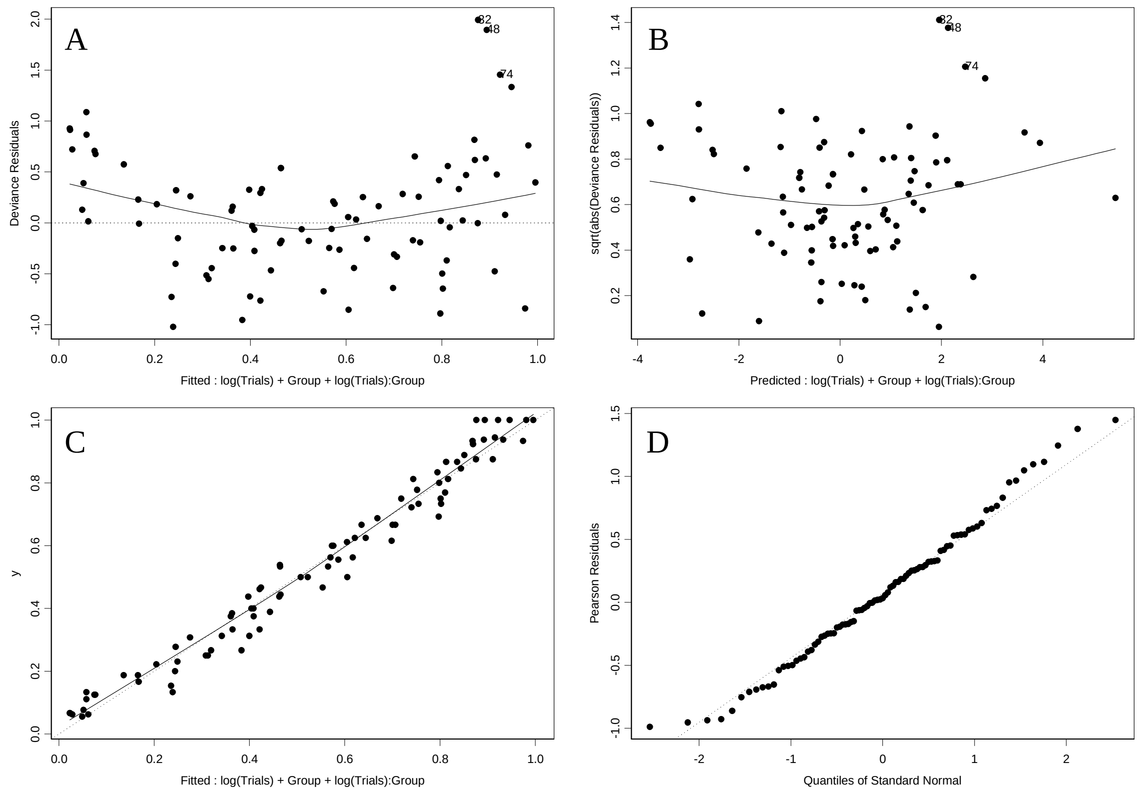


Figure V-30. Plots of the generalized linear model of `Proficient` (*y*) predicted by `log(Trials)`, `Group`, and `log(Trials):Group`.

The graphical displays (Figure V-31) for the model with the `probit` link function show no improvement in the systematic curvature in the residual plot (Figure V-31A) or the extreme observation on the normal quantile plot (Figure V-31B).

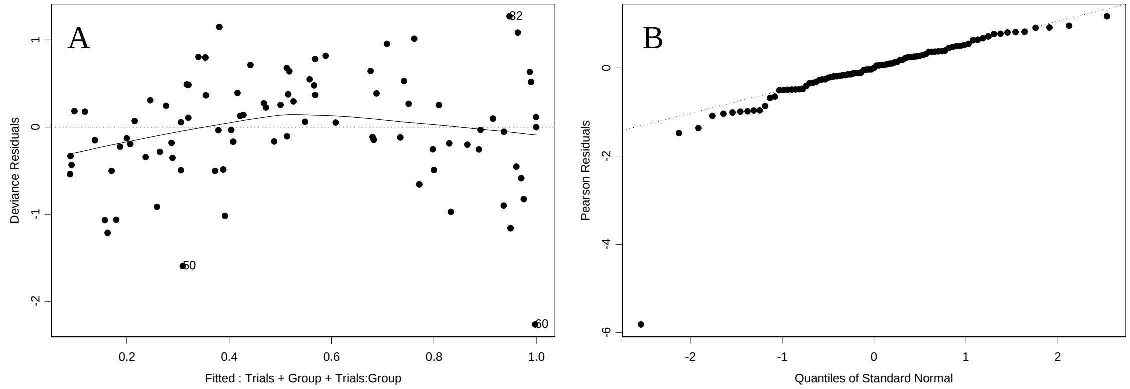


Figure V-31. Plots of the generalized linear model of `Proficient` (*y*) predicted by `Trials`, `Group`, and `Trials:Group`, using `probit` link function.

So far we have examined only linear relationships between the predictors and the proportion proficient. We now assess the validity of the linear assumption by fitting an additive model with relationships estimated by smoothing operations, and then comparing the linear fit. We first use the `gam` function to fit an additive model as follows:

```
> landing.gam.all <- gam(y ~ s(Trials) + Group + s(Trials):Group,
+ weight = Weight, family = binomial(logit), data = landing.task.df)
```

Indicating a predictor variable as an argument to the `s` function instructs `gam` to estimate the “smoothed” relationships with each predictor by using cubic B-splines. A summary of the fit is:

```
> summary(landing.gam.all)

Call: gam(formula = y ~ s(Trials) + Group + s(Trials):Group, family =
binomial(logit), data = landing.task.dfa, weights = Weight)
Deviance Residuals:
       Min         1Q      Median         3Q      Max
 -1.647109 -0.3010428 0.03344566 0.3292322 1.532508

(Dispersion Parameter for Binomial family taken to be 1 )

    Null Deviance: 551.8535 on 88 degrees of freedom

Residual Deviance: 23.71869 on 74.14184 degrees of freedom

Number of Local Scoring Iterations: 6

DF for Terms and Chi-squares for Nonparametric Effects
```

```
                Df Npar Df Npar Chisq     P(Chi)
    (Intercept)  1
      s(Trials)  1      2.9   6.495466 0.08111034
          Group  5
s(Trials):Group  5
```

Since the non-parametric tests do not inform us about the contribution of `Group` and the interaction term in the presence of a smooth of `Trials`, we fit two additional models that build on a base model: one with the `Group` variable and one with a smooth of the `Trials:Group` variable.

```
> landing.gam.trial <- gam(y ~ s(Trials), weight = Weight, family =
+ binomial(logit), data = landing.task.df)

> landing.gam.trial.group <- gam(y ~ s(Trials) + Group, weight =
+ Weight, family = binomial(logit), data = landing.task.df)

> landing.gam.all <- gam(y ~ s(Trials) + Group + s(Trials):Group,
+ weight = Weight, family = binomial(logit), data = landing.task.df)
```

We then produce the following analysis of deviance table:

```
> anova(landing.gam.trial, landing.gam.trial.group, landing.gam.all,
+ test = "Chisq")

Analysis of Deviance Table

Response: y

                                Terms Resid. Df Resid. Dev       Df
1                           s(Trials)  84.05436   160.0074
2                   s(Trials) + Group  79.12215    47.7865 4.932211
3 s(Trials) + Group + s(Trials):Group  74.14184    23.7187 4.980315

  Deviance              Test       Pr(Chi)
1
2 112.2209           + Group  0.0000000000
3  24.0678 +s(Trials):Group  0.0002068479
```

The indication is that `Group` is important in the model even with `Trials` included, and the interaction term, `Trials:Group`, is important even in the presence of both `Trials` and `Group`. Figure V-32 shows the graphical displays for the plots of the partial residuals (Figure V-32A) and the pointwise confidence intervals for the model that includes the `Trials` and `Group` variables and interaction term (Figure V-32B). The plots suggest a possible piecewise linear relationship for `Trials`.

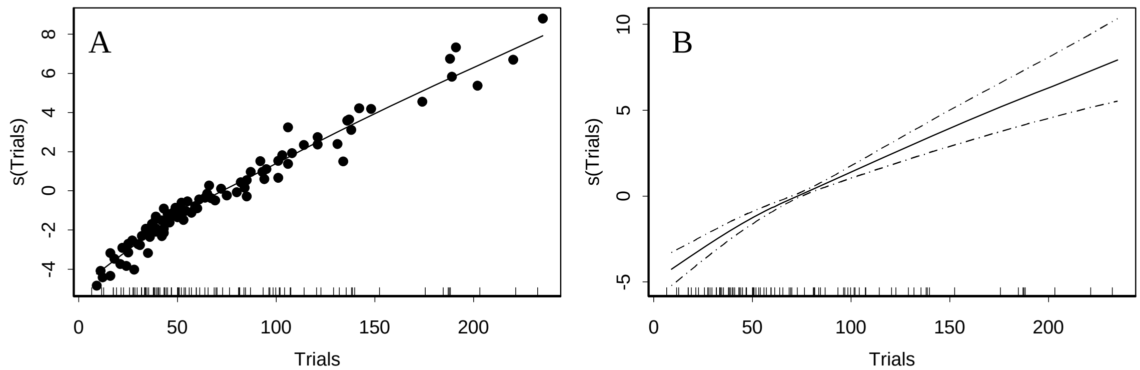


Figure V-32. The partial fits for the generalized additive logistic regression model of `Proficient` (*y*) with `Trials`, `Group`, and `Trials:Group` as predictors.

We next use the `anova` function to compare the linear fit `landing.glm.all` with the additive fit `landing.gam.all` to investigate whether if it may be worthwhile proceeding to develop a more complex model:

```
> anova(landing.glm.all, landing.gam.all, test = "Chisq")

Analysis of Deviance Table

Response: y

                                Terms Resid. Df Resid. Dev
1       Trials + Group + Trials:Group  77.00000   30.27233
2 s(Trials) + Group + s(Trials):Group  74.14184   23.71869

     Test       Df   Deviance    Pr(Chi)
1
2 1 vs. 2 2.858162   6.553645 0.07902184
```

From the previous table we see that the linear fit is more parsimonious. The effective degrees of freedom are 77 with the linear model and approximately 74 in the additive model with smooths. The residual deviance in the linear fit is not significantly higher than the residual deviance in the additive fit. In addition, with the linear fit, we can produce an analytical expression for the model, which cannot be done for an additive model with smooth fits. Given these considerations, we decide to use the linear fit to develop our subsequent isoperformance model. We have already shown that the `logit` link function provides the best fitting model. The only remaining question is whether to use the model with the log transformation of the predictor, `Trials`. While the variable transformation improves the appearance of the normal quantile plot by apparently resolving a possible extreme observation, in contrast, the systematic curvature in the

residual plots appears to be exacerbated and the plot of the absolute residuals against predicted values suggests the assumed variance function is less adequate with the transformation. Since we already demonstrated that none of the observations are excessively influential, we elect to use the model without the transformed predictor.

3. Reconnaissance Task Regression Model

Our initial logistic regression model relates the proportion proficient to the two predictor variables, `Trial` and `Group` and, as well as their interaction, `Trial:Group`. We fit the general linear model as follows:

```
> recon.glm.all <- glm(y ~ Trial + Group + Trial:Group, weight =
+ Weight, family = binomial(logit), data = recon.task.df)
```

The summary of the resulting fit:

```
> summary(recon.glm.all)

Call: glm(formula = y ~ Trial + Group + Trial:Group, family =
binomial(logit), data = recon.task.df, weights = Weight)
Deviance Residuals:
       Min          1Q      Median         3Q      Max
 -1.937115 -0.3732629 -0.05083308 0.4065998 1.222877

Coefficients:
                               Value Std. Error    t value
            (Intercept) -2.17731719 0.47230069 -4.6100233
                  Trial  0.07274783 0.02699650  2.6947134
        GroupCivil_inst  0.25482440 0.61709330  0.4129431
        GroupCivil_priv  0.75942388 0.71394647  1.0636986
     GroupPred_selectee -0.09795261 0.63980274 -0.1530981
          GroupT1_grad  0.82612780 0.63836094  1.2941390
         GroupT38_grad  0.89584150 0.62351721  1.4367551
   TrialGroupCivil_inst  0.06607400 0.04007844  1.6486172
   TrialGroupCivil_priv  0.04723454 0.05281837  0.8942824
TrialGroupPred_selectee  0.21569902 0.05315532  4.0579009
      TrialGroupT1_grad  0.09406102 0.04683788  2.0082255
     TrialGroupT38_grad  0.12591415 0.04650693  2.7074278

(Dispersion Parameter for Binomial family taken to be 1 )

    Null Deviance: 252.6496 on 46 degrees of freedom

Residual Deviance: 18.88778 on 35 degrees of freedom

Number of Fisher Scoring Iterations: 4
```

The test that all slopes are zero, G = 233.7618, DF = 11, and P-value < 0.001, indicates the model is adequate. The partial *t*-tests show that `Trial` is important even after

adjusting for `Group` and the interaction terms, but they provide less information on `Groups` or the interaction although these too seem to be important.

We next examine the bivariate relationships between the proportion proficient and each of the predictors by fitting a "null" model and then adding each of the terms, one at a time:

```
> recon.glm.null <- glm(y ~ 1, weight = Weight, family =
+ binomial(logit), data = recon.task.df)

> add1(recon.glm.null, ~ . + Trial + Group)

Single term additions

Model:
y ~ 1
       Df Sum of Sq      RSS       Cp
<none>              222.7744 232.4603
 Trial  1  121.4583 101.3161 120.6878
 Group  5   63.0341 159.7404 217.855
```

Using the `Cp` statistic to compare the models, `Trial` is clearly the best single variable to use in a linear model. However, to examine the contribution of `Group` and the interaction term to the full model, we produce an analysis of deviance for the sequential addition of each variable by using the `anova` function and specifying the chi-square test to evaluate for differences between models:

```
> anova(recon.glm.all, test = "Chisq")

Analysis of Deviance Table

Binomial model

Response: y

Terms added sequentially (first to last)
            Df Deviance Resid. Df Resid. Dev      Pr(Chi)
       NULL                    46   252.6496
      Trial  1 130.7356        45   121.9140 0.0000000000
      Group  5  81.9029        40    40.0111 0.0000000000
Trial:Group  5  21.1233        35    18.8878 0.0007677269
```

Here we see that `Group` is important after adjusting for `Trial`, and the interaction term, `Trial:Group`, is important after adjusting for both `Trial` and `Group`.

These statistical conclusions are subsequently verified by looking at graphical displays of the fitted values and residuals (Figure V-33). The plots indicate there may be

some problems with the model fit. A systemic curvature in the plot of deviance residuals versus the estimated proportion proficient (Figure V-33A) may be indicative of problems in the choice of link, the wrong scale for the predictor, or an omission of a quadratic term in the predictor. There does not appear to be large residuals suggesting outlying observations that might skew the analysis. The plot of the absolute residuals against predicted values (Figure V-33B) suggests the assumed variance function is adequate. The normal quantile plot (Figure V-33D) does not suggest any problem with normality.

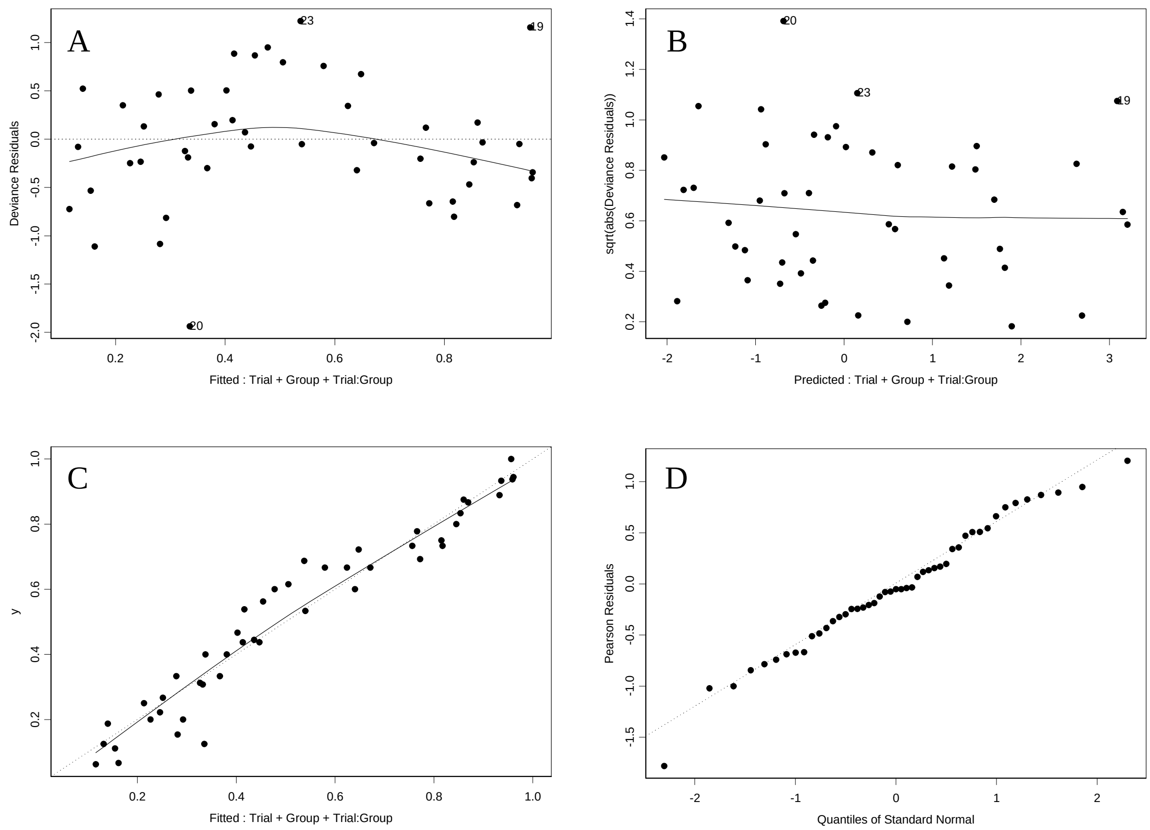


Figure V-33. Plots of the generalized linear model of `Proficient` (*y*) predicted by `Time`, `Group`, and `Trial:Group`.

We next examine several alternative models with different link functions and scaling of predictors to determine if the model fit can be improved. We fit models using the complementary log-log (`cloglog`) and `probit` link functions:

```
> recon.glm.cloglog <- glm(y ~ Trial + Group + Trial:Group, weight =
+ Weight, family = binomial(cloglog), data = recon.task.df)

> recon.glm.probit <- glm(y ~ Trial + Group + Trial:Group, weight =
+ Weight, family = binomial(probit), data = recon.task.df)
```

We also evaluate rescaling the predictor variable, `Trial`, using a log transformation:

```
> recon.glm.logtrial<- glm(y ~ log(Trial) + Group + log(Trial):Group,
+ weight = Weight, family = binomial(logit), data = recon.task.df)
```

We then use the `anova` function to compare these models and look at the graphical displays of the fitted values and residuals:

```
> anova(recon.glm.all, recon.glm.logtrial, recon.glm.cloglog,
+ recon.glm.probit)

Analysis of Deviance Table

Response: y

                                 Terms Resid. Df Resid. Dev
1           Trial + Group + Trial:Group        35   18.88778
2 log(Trial) + Group + log(Trial):Group        35   11.39551
3           Trial + Group + Trial:Group        35   26.10293
4           Trial + Group + Trial:Group        35   19.42642
```

The residual deviances suggest that the model using the rescaled predictor, `log(Trial)`, (#2) is the best, and the models using the `logit` (#1) and probit (#4) link functions appear comparable. Examining the graphical displays (Figure V-34) for the model with the rescaled predictor, we see that the introduction of the rescaled predictor resolves the systemic curvature in the plot of deviance residuals versus the estimated proportion proficient (Figure V-34A). It also appears to improve the fit between the observed and estimated proportion proficient (Figure V-34C), particularly for lower values of the response variable. However, the normal quantile plot (Figure V-34D) now suggests a light tailed distribution. Overall, our best fit appears to be obtained with the model using the rescaled predictor.

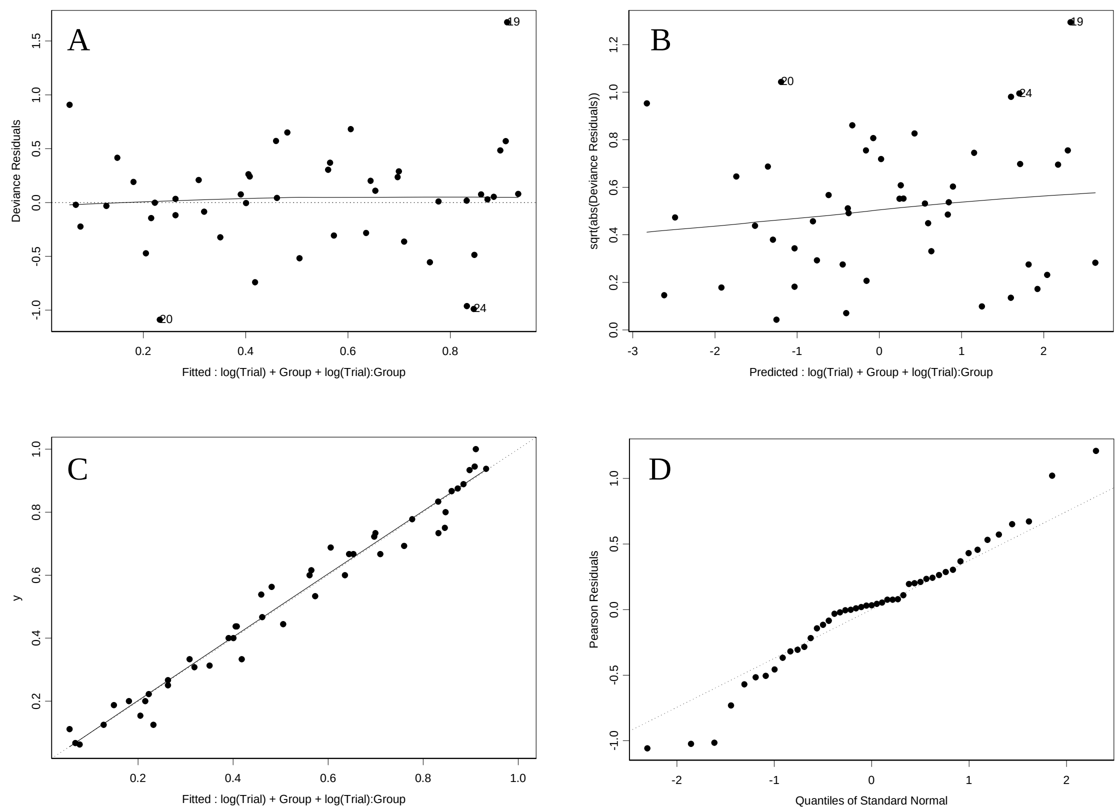


Figure V-34. Plots of the generalized linear model of `Proficient` (*y*) predicted by `log(Trial)`, `Group`, and `log(Trial):Group`

So far we have examined only linear relationships between the predictors and the proportion proficient. We now assess the validity of the linear assumption by fitting an additive model with relationships estimated by smoothing operations, and then comparing the linear fit. We first use the `gam` function to fit an additive model, indicating `Trial` as an argument to the `s` function, to estimate a "smoothed" relationship as follows:

```
> recon.gam.all <- gam(y ~ s(Trial) + Group + s(Trial):Group, weight =
+ Weight, family = binomial(logit), data = recon.task.df)
```

A summary of the fit is:

```
> summary(recon.gam.all)

Call: gam(formula = y ~ s(Trial) + Group + s(Trial):Group, family =
binomial(logit), data = recon.task.df, weights = Weight)
Deviance Residuals:
       Min         1Q     Median        3Q      Max
 -1.314452 -0.2170973 0.01017454 0.2258426 1.351369

(Dispersion Parameter for Binomial family taken to be 1 )
```

```
Null Deviance: 252.6496 on 46 degrees of freedom

Residual Deviance: 8.226457 on 32.05234 degrees of freedom

Number of Local Scoring Iterations: 5

DF for Terms and Chi-squares for Nonparametric Effects

               Df Npar Df Npar Chisq    P(Chi)
   (Intercept)  1
      s(Trial)  1     2.9   10.29804 0.0154536
         Group  5
s(Trial):Group  5
```

Since the non-parametric tests do not inform us about the contribution of `Group` and the interaction term in the presence of a smooth of `Trial`, we fit two additional models that build on a base model: one with the `Group` variable and one with a smooth of the `Trial:Group` variable.

```
> recon.gam.trial <- gam(y ~ s(Trial), weight = Weight, family =
+ binomial(logit), data = recon.task.df)

> recon.gam.trial.group <- gam(y ~ s(Trial) + Group, weight = Weight,
+ family = binomial(logit), data = recon.task.df)
```

We then produce the following analysis of deviance table:

```
> anova(recon.gam.trial, recon.gam.trial.group, recon.gam.all, test =
+ "Chisq")

Analysis of Deviance Table

Response: y

                            Terms Resid. Df Resid. Dev
1                        s(Trial)  41.91442   86.30809
2                s(Trial) + Group  37.01329   20.93000
3 s(Trial) + Group + s(Trial):Group  32.05234    8.22646

           Test       Df Deviance    Pr(Chi)

+Group          4.901128 65.37809 0.00000000
+s(Trial):Group 4.960953 12.70354 0.02565417
```

The indication is that `Group` is important in the model even with `Trial` included, and the interaction term, `Trial:Group`, is important even in the presence of both `Trial` and `Group`. Figure V-35 shows the graphical displays for the plots of the partial residuals (Figure V-35A) and the pointwise confidence intervals for the model that includes the `Trial` and `Group` variables and interaction term (Figure V-35B). The plots suggest a possible piecewise linear or logarithmic relationship for `Trial`.

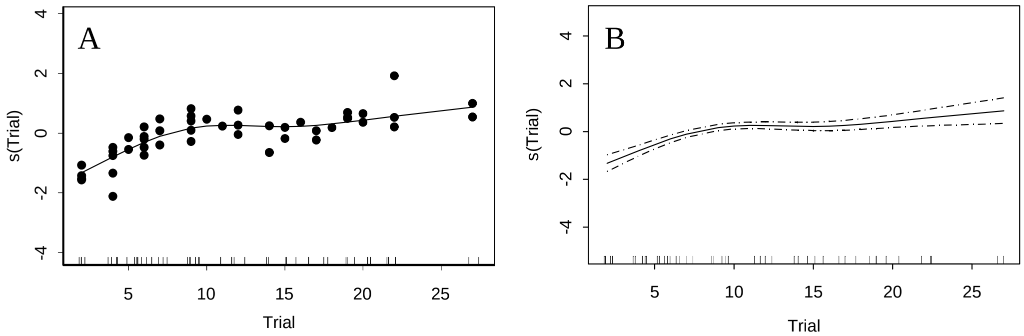


Figure V-35. The partial fits for the generalized additive logistic regression model of `Proficient` (*y*) with `Trial`, `Group`, and `Trial:Group` as predictors.

We next use the `anova` function to compare the linear fit `recon.glm.all` with the additive fit `recon.gam.all` to investigate whether if it may be worthwhile proceeding to develop a more complex model:

```
> anova(recon.glm.all, recon.gam.all, test = "Chisq")

Analysis of Deviance Table

Response: y

                                Terms Resid. Df Resid. Dev
1       Trial + Group + Trial:Group  35.00000   18.88778
2 s(Trial) + Group + s(Trial):Group  32.05234    8.22646

   Test        Df Deviance    Pr(Chi)

1 vs. 2 2.947661 10.66132 0.01306783
```

The additive fit appears to be significantly better than the linear fit, prompting us to next compare the linear fit using the rescaled predictor, `recon.glm.logtrial`, with the additive fit `recon,gam.all`:

```
> anova(recon.glm.logtrial, recon.gam.all, test = "Chisq")

Analysis of Deviance Table

Response: y

                                    Terms Resid. Df Resid. Dev
1 log(Trial) + Group + log(Trial):Group  35.00000   11.39551
2     s(Trial) + Group + s(Trial):Group  32.05234    8.22646

   Test        Df Deviance   Pr(Chi)

1 vs. 2 2.947661  3.16905 0.3575525
```

We see that the linear fit with the rescaled predictor is more parsimonious. The effective degrees of freedom are 35 with the linear model and approximately 32 in the additive model with smooth fit. The residual deviance in the linear fit is not significantly higher than the residual deviance in the additive fit. In addition, with the linear fit, we can produce an analytical expression for the model, which cannot be done for an additive model with smooth fits.

The summary of linear fit with the transformed predictor is as follows:

```
> summary(recon.glm.logtrial)

Call: glm(formula = y ~ log(Trial) + Group + log(Trial):Group, family =
binomial(logit), data = recon.task.df, weights = Weight)
Deviance Residuals:
       Min        1Q     Median        3Q      Max
 -1.088666 -0.253434 0.03289063 0.2514632 1.673539

Coefficients:
                                 Value Std. Error    t value
                 (Intercept) -3.0462734  0.7827949 -3.8915343
                  log(Trial)  0.8110230  0.3041620  2.6664181
             GroupCivil_inst -0.5714157  1.1022920 -0.5183887
             GroupCivil_priv -0.3495750  1.3793327 -0.2534378
          GroupPred_selectee -1.3587575  1.1204495 -1.2126897
                GroupT1_grad -0.9234459  1.2084209 -0.7641757
               GroupT38_grad  0.4257632  1.0244731  0.4155923
   log(Trial)GroupCivil_inst  0.6327380  0.4397222  1.4389494
   log(Trial)GroupCivil_priv  0.6603719  0.5845181  1.1297715
log(Trial)GroupPred_selectee  1.4636003  0.4768907  3.0690480
      log(Trial)GroupT1_grad  1.1904494  0.5161305  2.3064892
     log(Trial)GroupT38_grad  0.7886912  0.4278881  1.8432186

(Dispersion Parameter for Binomial family taken to be 1 )

    Null Deviance: 252.6496 on 46 degrees of freedom

Residual Deviance: 11.39551 on 35 degrees of freedom

Number of Fisher Scoring Iterations: 4
```

We have already examined the graphical displays for this model, and Tables V-11 and V-12 contain the S-Plus output for the residual analysis. The maximum standardized deviance residual is 1.83934 and the maximum standardized pearson residual is 1.34513, suggesting there are no outliers in the dataset. However, three observations (#37, 41, and 47) have leverages greater than twice $\bar{h}$, that is h_i > 0.51064, suggesting they are influential. Two of these observations correspond to the last observed participants within

the civil private pilot and cadet groups to become proficient during the 30 trials. However, D_{max} for the entire dataset is 0.12601, which is well below the threshold for concern, which is unity. Since there is no reason to doubt the validity of these influential observations, there is no justification for their removal (Montgomery, Peck, & Vining, 2006) and we decide to use the full linear fit with the rescaled predictor to develop our subsequent model.

Table V-11. Residuals for the reconnaissance task data.

Observation	Group	Trials	Observed Probability	Estimated Probability	Deviance Residuals	Pearson Residuals	h_{ii}	Cook's Distances
1	Pred_selectee	2	0.11111	0.05581	1.07971	1.21648	0.29409	0.05138
2	Pred_selectee	4	0.22222	0.22240	-0.00225	-0.00225	0.33340	0.00000
3	Pred_selectee	6	0.33333	0.41837	-0.83437	-0.82415	0.21234	0.01526
4	Pred_selectee	7	0.44444	0.50530	-0.56736	-0.56690	0.17021	0.00549
5	Pred_selectee	9	0.66667	0.64401	0.21771	0.21663	0.14119	0.00064
6	Pred_selectee	10	0.72222	0.69688	0.25487	0.25277	0.14361	0.00089
7	Pred_selectee	12	0.77778	0.77681	0.01074	0.01073	0.15914	0.00000
8	Pred_selectee	14	0.83333	0.83171	0.02018	0.02016	0.17428	0.00001
9	Pred_selectee	17	0.88889	0.88488	0.05939	0.05909	0.18556	0.00007
10	Pred_selectee	19	0.94444	0.90825	0.63210	0.58959	0.18618	0.00663
11	T38_grad	2	0.20000	0.18069	0.26817	0.27177	0.48852	0.00588
12	T38_grad	4	0.40000	0.40063	-0.00607	-0.00607	0.32503	0.00000
13	T38_grad	6	0.60000	0.56114	0.34078	0.33950	0.20200	0.00243
14	T38_grad	9	0.66667	0.70980	-0.39412	-0.39949	0.15112	0.00237
15	T38_grad	14	0.73333	0.83219	-1.04858	-1.11811	0.16030	0.01989
16	T38_grad	15	0.80000	0.84704	-0.53291	-0.55338	0.16330	0.00498
17	T38_grad	16	0.86667	0.85995	0.08273	0.08217	0.16590	0.00011
18	T38_grad	20	0.93333	0.89769	0.53107	0.50046	0.17168	0.00433
19	T38_grad	22	1.00000	0.91087	1.83934	1.33156	0.17216	0.03073
20	T1_grad	4	0.12500	0.23235	-1.44029	-1.34513	0.42866	0.11313
21	T1_grad	6	0.43750	0.40527	0.30828	0.30947	0.27989	0.00310
22	T1_grad	7	0.56250	0.48125	0.74232	0.74219	0.23193	0.01386
23	T1_grad	9	0.68750	0.60539	0.76223	0.75095	0.19922	0.01169
24	T1_grad	17	0.75000	0.84565	-1.16686	-1.25008	0.28236	0.05124
25	T1_grad	19	0.87500	0.87253	0.03535	0.03525	0.29217	0.00004
26	T1_grad	27	0.93750	0.93257	0.09419	0.09310	0.28576	0.00029
27	Civil_inst	2	0.06667	0.06806	-0.02560	-0.02552	0.30300	0.00002
28	Civil_inst	5	0.20000	0.21517	-0.16470	-0.16326	0.23317	0.00068
29	Civil_inst	6	0.26667	0.26293	0.03667	0.03672	0.19532	0.00003
30	Civil_inst	7	0.33333	0.30826	0.22837	0.23001	0.16442	0.00087
31	Civil_inst	9	0.40000	0.39045	0.08109	0.08121	0.12831	0.00008
32	Civil_inst	11	0.46667	0.46115	0.04573	0.04575	0.12204	0.00002
33	Civil_inst	15	0.53333	0.57251	-0.33367	-0.33479	0.16070	0.00179

Table V-12. Residuals for the reconnaissance task data (continued).

Observation	Group	Trials	Observed Probability	Estimated Probability	Deviance Residuals	Pearson Residuals	h_{ii}	Cook's Distances
34	Civil_inst	18	0.60000	0.63537	-0.31738	-0.31941	0.20612	0.00221
35	Civil_inst	19	0.66667	0.65325	0.12412	0.12373	0.22156	0.00036
36	Civil_inst	22	0.73333	0.69953	0.33701	0.33318	0.26536	0.00334
37	Civil_priv	4	0.15385	0.20488	-0.67568	-0.65275	0.51227	0.03729
38	Civil_priv	6	0.30769	0.31876	-0.10489	-0.10456	0.32861	0.00045
39	Civil_priv	9	0.53846	0.45937	0.65283	0.65427	0.23517	0.01097
40	Civil_priv	12	0.61538	0.56474	0.44025	0.43789	0.29274	0.00661
41	Civil_priv	22	0.69231	0.75993	-0.91359	-0.93993	0.63120	0.12601
42	Cadet	2	0.06250	0.07698	-0.28741	-0.27869	0.39214	0.00418
43	Cadet	4	0.12500	0.12764	-0.03787	-0.03776	0.29576	0.00005
44	Cadet	5	0.18750	0.14919	0.48370	0.49954	0.25857	0.00725
45	Cadet	12	0.25000	0.26290	-0.13175	-0.13104	0.20003	0.00036
46	Cadet	20	0.31250	0.35054	-0.39615	-0.39259	0.34016	0.00662
47	Cadet	27	0.43750	0.40775	0.34592	0.34712	0.51334	0.01059

G. BIBLIOGRAPHY

Archer, S., Headley, D., & Allender, L. (2003). Manpower, personnel, and training integration methods and tools. In H. R. Booher (Ed.), *Handbook of human systems integration* (chap. 11, pp. 379−431). New York, NY: John Wiley and Sons.

Barnes, M., & Beevis, D. (2003). Human systems measurements and trade-offs in system design. In H. R. Booher (Ed.), *Handbook of human systems integration* (chap. 8, pp. 233−263). New York, NY: John Wiley and Sons.

Beevis, D. (1996). Human functions and system functions. In: D. Beevis, P. Essens, & H. Schuffel (Eds.), *Improving function allocation for integrated systems design* (chap. 4, pp. 63–79) (Report No. CSER1AC SOAR 96-01). Wright-Patterson Air Force Base, OH: Crew Systems Ergonomics Information Analysis Center. (DTIC Accession No. A491093)

Bellman, R. E. (1961). *Adaptive control processes*. Princeton, NJ: Princeton University Press.

Blanchard, B. S., & Fabrycky, W. J. (2006). *Systems engineering and analysis* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Booher, H. R. (Ed.). (1990). *MANPRINT: an approach to systems integration*. New York, NY: Van Nostrand Reinhold.

Dahlman, C. J. (2007). *The cost of a military person-year: a method for computing savings from force reductions*. Santa Monica, CA: RAND.

Defense Acquisition University (2005). *Glossary of defense acquisition acronyms and terms* (12th Ed.). Ft. Belvoir, VA: Defense Acquisition University Press.

Defense Acquisition University (2009). Human systems integration. *Interim defense acquisition guidebook* (chap. 6). Ft. Belvoir, VA: Defense Acquisition University Press.

Department of Defense (1991). *Defense acquisition management policies and procedures* (DoDI 5000.2). Washington, DC: Author.

Department of Defense (2008). *Operation of the defense acquisition system* (DoDI 5000.02). Washington, DC: Author.

de Weck, O. L., & Jones, M. B. (2006). Isoperformance: analysis and design of complex systems with desired outcomes. *Systems Engineering*, 9(1), 45−61.

Hoffman, J. C., & Kamps, C. T. (2005, Spring). Future "manning" for unmanned aerial vehicles. *Air and Space Power Journal*, 31−37.

Hughes, T. P. (1998). *Rescuing Prometheus*. New York, NY: Vintage Books.

Jones, M. B. (2000). Isoperformance and personnel decisions. *Human Factors*, 42(2), 299−317.

Jones, M. B., & Kennedy, R. S. (1996). Isoperformance curves in applied physiology. *Human Factors*, 38(1), 167−182.

Jones, M. D., Kennedy, R. S, Kuntz, L. A., & Baltzey, D. R. (1987). Isoperformance: trading off selection, training, and equipment variations to maintain the same level of systems performance. In *Proceedings of the Human Factors Society 31st Annual Meeting 1987* (pp. 634−637). Santa Monica, CA: Human Factors Society.

Jones, M. B., Kennedy, R. S., & Stanney, K. M. (2004). Toward systematic control of cybersickness. *Presence*, 13(5), 589−600.

Jones, M. B., Kennedy, R. S., Turnage, J. J., Kuntz, L. A., & Jones, S. A. (1985). *Meta-analysis of human factors engineering studies comparing individual differences, practice effects and equipment design variations* (Report No. EOTR-86-7). Wright-Patterson Air Force Base, OH: Aeronautical Systems Division/PMRSB. (DTIC Accession No. A168790)

Jones, M. B., & Turnage, J. J., & Kennedy, R. S. (1993). *Isoperformance as a framework for human systems integration* (Report No. HSC-TR-93-001). Brooks Air Force Base, TX: Human Systems Center.

Kennedy, B. S., & Jones, M. B. (1992). *Simulating the impact of MPTS trade-off decisions by application of the isoperformance methodology*. Orlando, FL: Naval Training Systems Center. (DTIC Accession No. A274607)

Kennedy, R. S., Jones, M. B., & Baltzley, D. R. (1988). *Empirical demonstration of isoperformance methodology preparatory of an interactive expert computerized decision aid* Report No. ARI Research Note 88-93. Alexandria, VA: United States Army Research Institute for the Behavioral and Social Sciences. (DTIC Accession No. A202429)

Kennedy, R. S., Jones, M. B., & Baltzley, D. R. (1989). *Regions of isoperformance: an interactive software package for trading off training, personnel, and equipment* (Report No. AAMRL-TR-89-003). Wright-Patterson Air Force Base, OH: Harry G. Armstrong Aerospace Medical Research Laboratory. (DTIC Accession No. B136770)

Martin, E., Lyon, D. R., & Schreiber, B. T. (1998). Designing synthetic tasks for human factors research: an application to uninhabited air vehicles. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting* (pp. 123–127). Santa Monica, CA: Human Factors and Ergonomics Society.

Matthews, G., Davies, D. R., Westerman, S. J., & Stammers, R. B. (2000). *Human performance: cognition, stress, and individual differences*. Philadelphia, PA: Taylor and Francis.

McCracken, J. H., & Aldrich, T. B. (1984). *Analysis of selected LHX mission functions: implications for operator workload and system automation goals* (Technical Note ASI 479-024-84(b)). Fort Rucker, AL: Anacapa Sciences, Inc.

Meister, D. (1985). *Behavioral analysis and measurement methods*. New York, NY: John Wiley and Sons.

Mitchell, D. K. (2000). *Mental workload and ARL workload modeling tools* (Report No. ARL-TN-161). Aberdeen Proving Ground, MD: Army Research Laboratory.

Montgomery, D. C., Peck, E. A., & Vining, C. G. (2006). *Introduction to linear regression analysis* (4th ed.).Hoboken, NJ: John Wiley and Sons.

Nagy, J. E., Kalita, S. W., & Eaton, J. (2006). *U.S. Air Force unmanned aircraft systems performance analyses: Predator pilot front end analysis report* (Contract No. SPO700-98-D-4001). Brooks City-Base, TX: 311th Human Systems Wing Performance Enhancement Directorate.

Nelson, A., Schmitz, E. J., & Promisel, D. (1984). *The impact of personnel quality on STINGER weapon system performance* (Report No. ARI-TR-640). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC Accession No. ADA169235)

Office of the Undersecretary of Defense (Comptroller) (2010). *Military personnel composite standard pay and reimbursement rates*. Retrieved April 24, 2010, from DoD Comptroller Web site: http://comptroller.defense.gov/rates/index.html

Pew, R. W., & Mavor, A. S. (2007). *Human-system integration in the system development process: a new look*. Washington, DC: National Academies Press.

Pew, R. W. (2008). More than 50 years of history and accomplishments in human performance model development. *Human Factors*, 50(3), 489−496.

Pohl, E. (2008). System effectiveness. In G. S. Parnell, P. J. Driscoll, & D. L. Henderson (Eds.), *Decision making in systems engineering and management*. Hoboken, NJ: John Wiley and Sons Inc.

Rardin, R. L. (1998). *Optimization in operations research*. Upper Saddle River, NJ: Prentice Hall.

Schreiber, B. T., Lyon, D. R., Martin, E. L., & Confer, H. A. (2002). *Impact of prior flight experience on learning Predator UAV operator skills* (Report No. AFRL-HE-AZ-TR-2002-0026). Mesa, AZ: Air Force Research Laboratory Human Effectiveness Directorate, Warfighter Training Research Division. (DTIC Accession No. A401588)

Simon, H. A. (1996). *The sciences of the artificial* (3rd ed.). Cambridge, MA: MIT Press.

Weisz, J. D. (1967). *Systems analysis: human factors research and development contributions* (Technical Note 6-67). Aberdeen Proving Ground, MD: Human Engineering Laboratory. (DTIC Accession No. AD818566)

Weisz, J. D. (1968). *Systems analysis: manpower resource/system design integration methodology* (Technical Note 9-68). Aberdeen Proving Ground, MD: Human Engineering Laboratory. (DTIC Accession No. AD675481)

Wickens, C. D., Bellenkes, A. H., & Kramer, A. F. (1995). *Expertise and attention control in pilot visual scanning* (Report No. UIUC-BI-HPP-95-5). Urbana, IL: University of Illinois Beckman Institute of Advanced Science and Technology.

VI. HUMAN SYSTEMS INTEGRATION DOMAIN TRADEOFFS IN NON-TECHNICAL SYSTEMS – IMPROVING SOLDIER BASIC COMBAT TRAINING

> The bed is a bundle of paradoxes: we go to it with reluctance, yet we quit it with regret; we make up our minds every night to leave it early, but we make up our bodies every morning to keep it late (Colton, 1837, p. 164).

A. INTRODUCTION

Is human systems integration (HSI) mainly an adjunct or enabler to the systems engineering and management process as asserted by Booher (2003), and more recently Deal (2007)? Or is it primarily the incorporation of domain concerns that are relevant to system design within human factors engineering activities as suggested by Pew and Mavor (2007)? If we examine Department of Defense (DoD) acquisition guidance, HSI is championed by an acquisition program manager and executed within the framework of systems engineering (Defense Acquisition University [DAU], 2009; DoD, 2008). However, in juxtaposition to this view is the reality that Defense Department strategic planning guidance directs the conduct of analyses across HSI domains when evaluating both materiel and non-materiel options for satisfying identified functional needs (DoD, 2009). Decisions to pursue non-materiel solutions do not involve the Defense Department acquisition system or the systems engineering framework, thereby necessitating that HSI considerations be managed independent of the system acquisition process.

Such contradictions drive debate (at times, intense) about the very nature of HSI, which frustrates subsequent efforts to expound upon the concept. In response, our earlier case history of the Human Performance Integration Directorate (Chapter II) used soft systems methodology to first define HSI as a set of purposeful activities occurring within a human performance-generating operation. One of the main activities identified in the study, conducting human performance/HSI analyses, involved the decomposition of human performance criteria into multi-domain solution sets as a precondition for scientific research and engineering potential solutions.

In Chapter V, we built upon this *Weltanschauung* of HSI by showing that human performance could be described in terms of human factors engineering, manpower, personnel, and training domain solution sets. Additionally, these same domain sets were shown to be determinants of reliability and the safety domain of HSI. We now wish to extend our HSI model to include the personnel survivability domain, at least as it pertains to issues of physical and mental fatigue (Zigler & Weiss, 2003). We do so by examining a problem situation concerning a human performance-generating operation involving non-technical systems—Basic Combat Training—as a process for engineering human weapon systems (Deuster et al., 2007; Tvaryanas, Brown & Miller, 2009), which is to say, U.S. Army Soldiers. We deliberately chose this use case to gain insight into the applicability of our emerging HSI model in a non-materiel context. Is HSI an adjunct to systems engineering of technical systems or a more general systems approach to problematic situations involving human performance? We start by considering the narrower question of the role of fatigue, and hence, the survivability domain, as a determinant of Soldier performance during Basic Combat Training.

B. EFFECTS OF SLEEP ON TRAINING EFFECTIVENESS IN SOLDIERS AT FORT LEONARD WOOD

1. Statement of the Problem

Military training regimes often include some degree of sleep deprivation, whether it is by design or unintentional. Several studies have demonstrated that sleep deprivation is prevalent in military training and education programs. For example, Killgore and colleagues (2008), using actigraphy to assess sleep in Soldiers attending military training at the Noncommissioned Officer Academy and the Warrant Officer Candidate School, reported Soldiers obtained an average of 5.8 hours of sleep per night. Miller and colleagues (2008), reporting on the preliminary results of a 4-year longitudinal study of sleep in U.S Military Academy (USMA) cadets based on actigraphy data, found that cadets averaged 5.4 hours of sleep per night. This is substantially less than the approximately eight hours of sleep per night required by healthy adults to maintain cognitive effectiveness (Anch et al., 1988). Additionally, this is more than two hours less sleep per night than cadets stated receiving prior to arriving at the USMA (Miller, 2005).

It is also important to recognize that military recruits are adolescents or young adults in their late teens and early twenties. Biologically driven sleep-wake patterns in this age group differ from those of more mature adults, with delayed bedtimes, later awakenings, and longer sleep periods (i.e., on the order of 0.5 to 1.25 more hours of sleep per night) (Carskadon et al., 1997, 1998; Wolfson & Carskadon, 2003). Thus, the general population of military recruits may actually require from 8.5 to 9.25 hours of sleep per night for optimal performance (Miller & Shattuck, 2005).

Chronic sleep deprivation from multiple nights of less than eight hours of sleep will cause sleep debt and fatigue. A vast body of research has shown that the effects of fatigue include decreased vigilance, adverse mood changes, perceptual and cognitive decrements (Krueger, 1990; Belenky et al., 2003; Van Dongen et al., 2003), impaired judgment and increased risk taking (Killgore, Balkin, & Wesensten, 2006), and even decreased marksmanship (Tharion, Shukitt-Hale, & Lieberman, 2003; McLellan et al., 2005). Contrary to popular opinion in the military, research has shown that motivation can only partially compensate for the adverse effects of sleep deprivation (Pigeau, Angus, & O'Neil, 1995).

Of particular relevance to military training, the ability of individuals to learn and retain information is reduced by sleep deprivation (literature summarized in Miller, Matsangas, & Shattuck, 2007). For example, Graham (2000) reports that learning curves drop dramatically for adolescents obtaining 4–6 hours of sleep relative to those obtaining eight hours per night. In the military training environment, Andrews (2004) conducted a retrospective comparison of the academic performance of Navy recruits before and after the training command leadership changed the sleep regime from six to eight hours per night. It was observed that recruits who received eight hours of sleep per night scored on average 11% higher than their counterparts who received only six hours of sleep, although Andrews was unable to discount the impact of other, concurrent changes at the training command. In contrast, Baldus (2002) collected actigraphic data on 31 Navy recruits at the same training command who were all assigned to two sleep conditions (9:00 p.m. to 5:00 a.m. and 10:00 p.m. to 6:00 a.m.) in a cross-over study design. It was

shown that recruits obtained an additional 22 minutes of sleep when on the 1-hour phase-delayed sleep schedule, but no attempt was made to correlate this observation with measures of recruit performance.

However, Killgore and colleagues (2008), evaluating the effectiveness of actigraphy as a predictor of cognitive performance, found significant positive correlations between Soldier academic exam scores in six military education programs (i.e., programs of instruction at the Noncommissioned Officer Academy and Warrant Officer Candidate School at Fort Rucker, AL) and the following sleep indices: average hours of sleep per night and hours slept in the 24 and 48 hour periods preceding an exam. They also report that the average amount of sleep obtained by Soldiers accounted for approximately 40% of the variance in exam scores—a finding that underscores the impact of fatigue on learning and memory. A similar result was reported by Trickel and colleagues (2000) who found that sleep habits accounted for most of the variance in the academic performance of freshman college students.

Physical health is an equally important concern in military recruit populations, particularly because the close living conditions are conducive to the spread of communicable disease. Individual physical health, and in turn, public health, also depends on individuals receiving adequate amounts of sleep. Research has shown that disturbances of sleep-wake homeostasis are accompanied by alterations in the immunological, neuroendocrine, and thermoregulatory functions of the body, and hence, contribute to pathological processes such as infectious disease (Moldofsky, 1995). Lange and colleagues (2003) also report that sleep enhances antibody production and the immune response to vaccination. Besides illness, sleep deprivation threatens health by increasing the risk for injuries resulting from accidents. For example, Thorne and colleagues (1992) demonstrated that accidents increase progressively as sleep duration decreases to 7, 5, and 3 hours per night over a period of one week.

Scientific literature suggests there is a high prevalence of fatigue in military recruits, which has important implications for Soldier training, health, and safety. Well-controlled laboratory experiments have demonstrated a convincing dose-response relationship between sleep deprivation and degraded cognitive performance (Belenky et

al., 2003; Driskell, Hughes, Willis, Cannon-Bowers, & Salas, 1991; Driskell & Salas, 1996; Hursh & Bell, 2001; Van Dongen et al., 2003) (as discussed in Miller, Matsangas, & Shattuck, 2007). However, the design of prior studies of fatigue in military training environments has been primarily descriptive in nature, limited to correlations between sleep and academic test performance, and many of the recommendations for follow-on research have yet to be followed. The only field study to directly examine the effect of a phase-delayed sleep scheduling intervention in the military training environment (Baldus, 2002) did not include any assessment of performance outcomes. Thus, whether designing schedules to minimize fatigue would have a direct effect on outcomes in the military training environment remains an open question.

The scarcity of information on the benefit of sleep scheduling interventions for military training is regrettable because it is the sort of evidence that senior decision-makers require if they are to support fatigue-sensitive revisions to training regimes. If sleep scheduling is found to have a significant effect on overall training effectiveness and recruit attrition, health, and safety, then two options become available for the military training community:

- Performance thresholds of achievement for basic military training can be increased while maintaining the present length of training (optimizing training effectiveness), or
- Thresholds of achievement can be maintained and the length of training decreased (optimizing training efficiency).

Preliminary evidence suggests that sleep, and conversely fatigue, may account for nearly half the variability in academic performance during military training (Killgore et al., 2008). Additionally, implementing a phase-delayed sleep scheduling intervention during military training appears to result in measurable increases in total daily sleep (Baldus, 2002). Collectively, these observations suggest that sleep scheduling is a potentially powerful lever for manipulating the performance of military training programs—and one that is immediately within our grasp without making a significant investment in new technologies. Since training is a potential bottleneck in meeting wartime manpower needs as well as a recurring life-cycle cost for all weapon systems, even a more modest 10%

improvement in trainee performance as suggested by Andrews (2004) is significant when one considers the cumulative impact across military training programs.

This study attempts to contribute to the knowledge base by exploring the influence of sleep scheduling in the Basic Combat Training environment on Soldiers' achievement of entry-level standards and combat skills. This study examines the direct effect of sleep scheduling on motivation and mood state and training, health, and safety outcomes while controlling for such individual differences as sleep habits, personality, and personnel aptitudes.

2. Purpose of the Study

The purpose of this study is to examine the effect of alterations in the timing of sleep within the circadian cycle on the amount of total nightly sleep and its influence on various indicators of mood and performance of U.S. Army Soldiers attending Basic Combat Training at Fort Leonard Wood, Missouri. The study design compares Soldiers assigned to one of two training companies: a company using the standard Basic Combat Training sleep regimen (i.e., sleep period 8:30 p.m. to 4:30 a.m.) or a company using a phase-delayed sleep regimen (i.e., sleep period 11:00 p.m. to 7:00 a.m.), the latter being more in line with the biologically driven sleep-wake patterns of adolescents.

To account for some of the myriad factors that are assumed to play a role in daytime functioning, a number of factors are selected as control variables or covariates (Table VI-1). These control variables include background information about each Soldier (e.g., age, sex, caffeine and tobacco habits, prior experience with firearms, etc.) and information about their sleep habits, personality, resilience, and personnel aptitudes. The inclusion of these individual characteristics is important to this study because we predict that sleep timing will have a small, but measurable influence on daytime functioning even after controlling for the contributions of the usual variables thought to affect mood state and performance.

Table VI-1. Summary of study variables.

Independent variables	Dependent variables
Age	Attrition
Caffeine and tobacco habits	Basic rifle marksmanship
Personality	Mood state
Personnel aptitude	Physical fitness
Prior experience with firearms	
Resilience	
Sex	
Sleep habits	
Sleep schedule	

Consequently, at weekly intervals, Soldiers are asked to identify their mood state over the prior week of training. Mood state is defined by six general factors identified in the Profile of Mood States (POMS) (McNair, Lorr, & Droppleman, 1981). These six factors are tension-anxiety, depression-dejection, anger-hostility, vigor-activity, fatigue-inertia, and confusion-bewilderment. These six factors can also be aggregated into a total mood disorder score. The study primarily examines three major performance outcomes of concern to the military training organization: attrition, basic rifle marksmanship, and physical fitness.

3. Theoretical Perspective and HSI Model Elaboration

In formulating a theoretical perspective for considering the survivability domain of HSI in concert with the manpower, personnel, training, and safety domains within a systems context, fatigue models provide a useful prototype. Besides the typical survivability characteristics of susceptibility, vulnerability, and recoverability, personnel survivability includes issues related to physical and mental fatigue (Zigler & Weiss, 2003). To that end, the Defense Department has long pursued applied research concerning fatigue in military operations and has developed several fatigue models. One of these models, known as the Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE)

Model, has achieved relatively wide acceptance and has seen practical application within the Fatigue Avoidance Scheduling Tool (FAST) (Hursh et al., 2004).

The SAFTE model is shown in Figure VI-1 using a system dynamics modeling stock and flow diagram. The conceptual architecture of the SAFTE model centers on a sleep reservoir, representing sleep-dependent processes that govern the capacity to perform cognitive work. Using the language of system dynamics modeling, the stock of this reservoir is cognitive work capacity. Sleep is a replenishing flow into the reservoir, while wakefulness is a depleting flow out of the reservoir. Replenishment, in terms of sleep accumulation, is determined by information about the time-of-day of sleep, reservoir level (i.e., sleep debt), and sleep quality (i.e., sleep fragmentation). The system modeled in Figure VI-1 provides output in terms of performance effectiveness, which is simultaneously modulated by circadian effects and the level of the reservoir (Hursh et al., 2004).

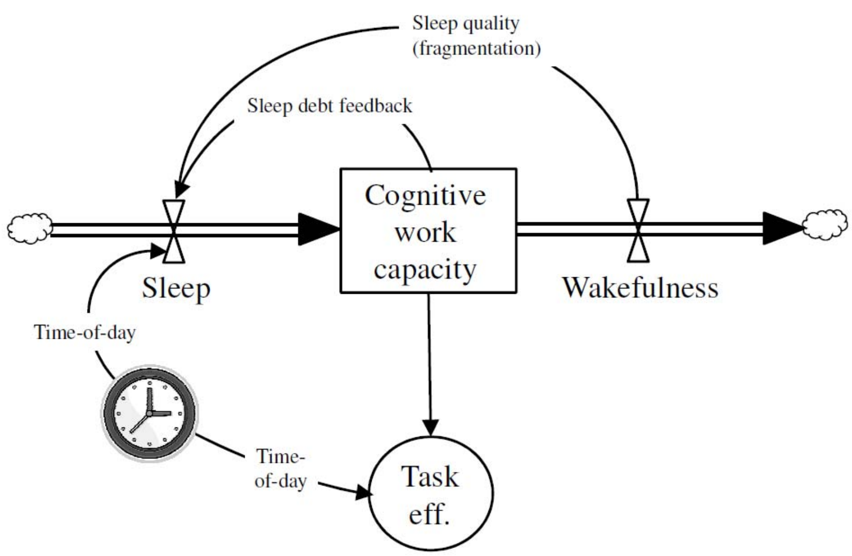


Figure VI-1. Stock and flow diagram of the SAFTE model.

The SAFTE model has been shown to predict changes in cognitive capacity as measured by standard laboratory tests of cognitive performance with reported coefficients

of determination ranging from 89%-94%. It is presumed that these cognitive tasks measure changes in the fundamental capacity to perform a variety of real world tasks that rely on such cognitive skills as discrimination, reaction time, mental processing, reasoning, and language comprehension and production. Although specific military tasks may vary in their reliance on these skills, Hursh and colleagues assert that it is reasonable to assume that changes in military task performance will correlate with changes in the underlying cognitive effectiveness. Hence, there is an expected direct relationship between measured changes in cognitive effectiveness and military task performance. Based on this reasoning, the SAFTE model can be used to predict variations in any task or component of a task, given appropriate data, using the following expression for generalized cognitive task effectiveness:

$$TE = A \cdot \left(\frac{R_t}{R_c} \right) + B + C_t + I \qquad (1)$$

where A is the linear component slope, $\frac{R_t}{R_c}$ is the reservoir level at time t expressed as a proportion of capacity, B is the linear component intercept, I is the transient sleep inertia term, and C_t is computed from the circadian process as follows:

$$C_t = C_1 \cdot \left[\cos\left(\frac{2\pi \cdot (T - p)}{24} \right) + C_2 \cdot \cos\left(\frac{4\pi \cdot (T - p - p')}{24} \right) \right] \qquad (2)$$

where T is the time of day in hours, p is the time of the peak of the 24-hour circadian rhythm, p' is the relative time of the 12-hour peak, and C_1 and C_2 are the 24- and 12-hour circadian weighting factors, respectively (Hursh et al., 2004).

Cognitive task effectiveness, as calculated by Equation 1, is the level of performance, expressed as a percent of some baseline. This construct can be generally related to the personnel and training domains of HSI through the latter's combined contribution to defining some performance baseline. A convenient and ubiquitous means for considering the personnel and training domains as determinants of performance is the power law of practice (Newell & Rosenbloom, 1981). This empirical regularity relates the personnel and training domain as determinants of performance:

$$P = a + b \cdot \left(N + E\right)^{-r} \quad (3)$$

where P is the time taken to perform a task, a is the asymptote or highest level of performance obtainable, b is the performance on the first trial, N is the amount of practice in terms of trials, E is the transfer from prior experience or learning required to attain entry level performance, and r is a learning rate parameter. With respect to the training domain, the quantity of training is directly reflected in the N variable and factors impacting training effectiveness are reflected by the value of the r parameter. In terms of the HSI personnel domain, to the extent that an individual's experience with prior tasks is similar to the target task, positive transfer occurs (Wickens & Hollands, 2000) as captured by the variable E. Also, since aptitude tests predict proficiency on various tasks and propensity for a variety of types of learning (Matthews, Davies, Westerman, & Stammers, 2000), individual aptitudes will influence the values for the a, b, and r parameters.

While the power law of practice is generally viewed as associated with perceptual-motor skills, it appears to hold for practice learning of all kinds. The law shows up everywhere in psychological behavior and cannot be easily circumscribed as applying to only some part of human operations (Newell & Rosenbloom, 1981). Overall then, the power law provides a general construct for modeling baseline performance in terms of the training and personnel domains of HSI and their interaction. Furthermore, it is a relatively simple matter to adjust baseline performance for circadian effects and level of the sleep reservoir as follows:

$$P' = TE \cdot P = \left[A \cdot \left(\frac{R_t}{R_c} \right) + B + C_t + I \right] \cdot \left[a + b \cdot \left(N + E\right)^{-r} \right] \quad (4)$$

where P' is the adjusted task performance and the other factors in Equation 4 are as previously described for the SAFTE model and power law of practice. In so doing, we have defined a performance solution set in terms of the personnel, training, and survivability domains of HSI.

Although the power law provides a simple mathematical construct that is easily modified to account for fatigue-related survivability domain considerations, further

elaboration is required if we are to examine the survivability domain within a broader systems context. Based on the structure of the SAFTE model, the reservoir or stock of cognitive work capability, shown in Figure VI-1, will reach a time-averaged equilibrium state provided an individual remains on a constant schedule (Hursh et al., 2004). Additionally, our stock and flow diagram of the SAFTE model shows that sleep accumulation is dependent on information regarding "sleep quality," which is modeled as the continuity, or conversely, fragmentation of sleep. The software implementation of the SAFTE model, the Fatigue Avoidance Scheduling Tool (FAST), addresses sleep quality in terms of the sleep environment and the average number of interruptions to sleep expected in that environment. The FAST software provides the following ordinal scale for describing sleep environments:

- *Excellent*: 0 interruptions per hour
- *Good*: 1-2 interruptions per hour
- *Fair*: 3-5 interruptions per hour
- *Poor*: 6 or more interruptions per hour

These values are equated to 60, 50, 40, and 30 minutes of effective sleep per hour, respectively.

Given the implications of the SAFTE model structure, it is clear that two classes of variables must be considered: schedule and sleep environment. The schedule determines the timing and duration of sleep and wakefulness, and in conjunction with sleep quality, determines the equilibrium state of the reservoir. In principle, the equilibrium state of the reservoir correlates inversely to the degree to which an individual is fatigued, the latter being a direct concern of the survivability domain of HSI. Likewise, the sleep environment is a determinant of sleep quality, which modulates sleep accumulation, and in turn, the equilibrium state of the reservoir. Since the sleep environment is shaped by the physical environment of sleeping or berthing areas (e.g., adequate space, temperature and lighting control, and noise attenuation), it is a direct consideration of the habitability domain of HSI. Consequently, the habitability domain, in terms of sleep environment and sleep quality, is a determinant of the survivability domain.

In the application of the synthesis of the power law and the SAFTE modeling framework to our emerging conceptualization of HSI, two new major determinants of skilled performance, besides training and personnel issues, are defined as follows:

- *Schedule* is a predetermined, recurring cycle of periods of wakefulness and sleep that is established by an organization for its workforce.
- *Sleep environment* is a physical space or berthing area that is designed or organizationally designated for sleeping, the adequacy of which is described in terms of sleep quality.

With these specific definitions, we can now operationalize the HSI survivability domain in terms of the variable, "schedule," and the habitability domain in terms of the variable, "sleep environment." The decision to describe the survivability domain using the term, schedule, reflects the predominant role of a schedule in determining the equilibrium state of the reservoir, and in turn, fatigue. Additionally, organizational planners and decision makers are likely to appreciate schedules as being within their purview but would struggle with the notion of setting fatigue levels. Furthermore, decision makers often must consider external, real-world demands or requirements to perform and available manpower resources when designing or choosing a schedule—the implication being that the manpower domain of HSI also influences the survivability domain, although modeling this concept is outside the scope of the current discussion.

Now it is possible to consider both the survivability and habitability domains of HSI within a broader system context using the *Weltanschauung* provided by the isoreliability construct developed in Chapter V, albeit with some modification. Previously, we classified a hypothetical system operator as being in one of two states, proficient or not proficient, if their performance, P, met or exceeded some *a priori* performance criterion, P_{ref}. We now reclassify our system operator based on whether $P' \geq P_{\text{ref}}$ where $P' = TE \cdot P$ and TE is a function of schedule and sleep environment. In other words, given some training time, x_1, personnel aptitude measure, x_2, schedule of work and sleep, x_3, and sleep environment, x_4, we focus on the operator being in one of two states:

- proficient $\left(X\left(x_1, x_2, x_3, x_4\right) = 1\right)$ and
- not proficient $\left(X\left(x_1, x_2, x_3, x_4\right) = 0\right)$.

Now suppose we have $N\left(0, x_2, x_3, x_4\right)$ initial trainees and we define the number of proficient graduating trainees after some period of training, x_1, as $N_s\left(x_1, x_2, x_3, x_4\right)$. Consequently, the *human reliability* can now be expressed in terms of training time, aptitude, schedule, and sleep environment:

$$R\left(x_1, x_2, x_3, x_4\right) = \frac{E\left[N_s\left(x_1, x_2, x_3, x_4\right)\right]}{N\left(0, x_2, x_3, x_4\right)} = \pi_i$$

where π_i is simply the probability the i^{th} trainee is proficient. Assuming a logistic regression model as we did in Chapter V, we can factor in our assurance level, α, and express our human reliability function as follows:

$$R\left(x_1, x_2, x_3, x_4\right) = \frac{1}{1 + \exp\left(-\boldsymbol{x}\hat{\boldsymbol{\beta}} + z_\alpha \sqrt{\operatorname{Var}\left(\boldsymbol{x}\hat{\boldsymbol{\beta}}\right)}\right)}$$

Once fitted to data, human reliability is modeled as the probability of satisfactory operator performance, or conversely, the probability of unsatisfactory operator performance. Given the latter, we can make statistical inferences relevant to the safety domain of HSI. All of this then allows us to extend our basic systems integration model for HSI, as proposed in Chapter V (see Figure V-5), so that it now includes the survivability and habitability domains as shown in Figure VI-2.

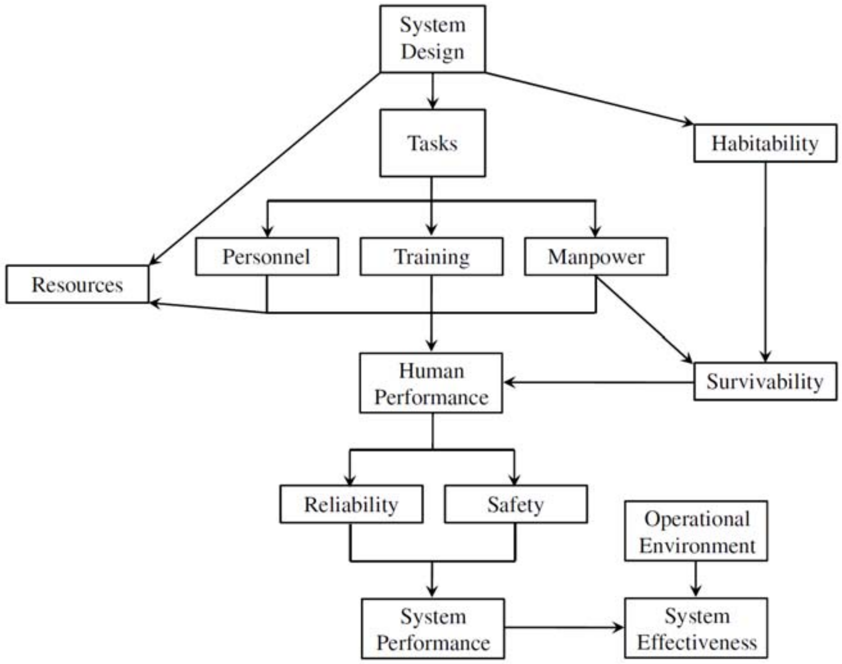


Figure VI-2. Manpower, personnel, training, safety, survivability, and habitability domains within a system structure.

In the application of our expanded HSI model to this study of Basic Combat Training, we have a somewhat simpler constrained problem. Specifically, two of the classes of variables discussed in the reliability formulation, training and sleep environment, are fixed as follows:

- *Training*: The duration and content of the basic military training program are fixed based on the Army's official program of instruction (POI)—that is, $x_1 = x_{\text{POI}}$.
- *Habitability*: The choice of sleep environment is dictated by the existing training barracks—that is, $x_4 = x_{Barracks}$.

Therefore, the following statement represents the underlying logic for designing and conducting this study. If we design a schedule so the timing of sleep-wake periods improves the overall equilibrium state of Soldiers' reservoirs—and consequently cognitive task effectiveness—then 1) individual Soldier task performance should improve resulting in a greater proportion of recruits who meet specified performance criteria, and

2) this effect should be greater for those Soldiers with lower personnel aptitudes as their performance margin relative to the specified performance criteria is expected to be smaller.

The predicted relationship between personnel aptitude, schedule, and their interaction and the outcome, proportion of the population that is proficient, is illustrated in Figure VI-3. As shown, Schedule 2 results in a more favorable equilibrium state of Soldiers' reservoirs than Schedule 1, which is to say that Schedule 2 is more complementary to Soldiers' natural circadian cycles. Hence, Schedule 2 is more effective overall, but it is particularly beneficial for recruits on the lower end of the personnel aptitude spectrum.

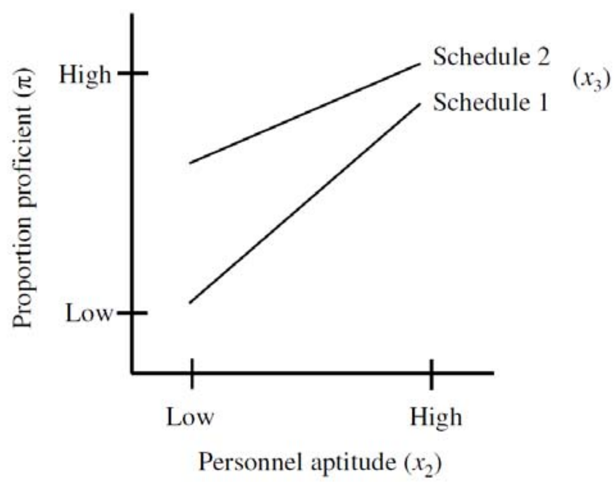


Figure VI-3. Hypothetical interactive effects of aptitude and two training schedules on learning outcomes.

It is worth noting that if we replaced the word "schedule" with "treatment" in Figure VI-3, we would have the depiction of an ordinal aptitude treatment interaction (ATI) as described in ATI theory (Whitener, 1989), which is yet another *Weltanschauung* for considering this study. The underlying premise of ATI theory is that learning, and subsequent performance, is higher when the learning method, or treatment, capitalizes on an individual's cognitive aptitudes (Snow, 1978). In a twist on ATI theory, this study

involves no change in learning methods *per se*, but rather, the treatment changes the relative availability of cognitive resources. Again, the underling logic for this study would suggest that if a schedule enhances cognitive resources, then 1) this should be manifest by increased performance on learning tasks, and 2) performance enhancements should be greater for those with less aptitude given their overall higher demand for cognitive resources during training.

4. Study Hypotheses

The following hypotheses guide this study:

H_1: Participants on the modified, phase-delayed sleep schedule will obtain more daily sleep than participants following the standard Basic Combat Training schedule.

H_2: Participants on the modified sleep schedule will have less decrement in mood state than participants following the standard Basic Combat Training sleep schedule.

H_3: Participants on the modified sleep schedule will exhibit greater improvement in basic rifle marksmanship scores than participants following the standard Basic Combat Training sleep schedule.

H_4: Participants on the modified sleep schedule will exhibit greater improvement in physical fitness scores than participants following the standard Basic Combat Training sleep schedule.

H_5: The likelihood of participants on the modified sleep schedule reporting occupationally significant fatigue will be lower than that for participants following the standard Basic Combat Training sleep schedule.

H_6: The likelihood of participants on the modified sleep schedule reporting poor sleep quality will be lower than that for participants following the standard Basic Combat Training sleep schedule.

H_7: The likelihood of participants on the modified sleep schedule attriting from training will be lower than that for participants following the standard Basic Combat Training sleep schedule.

5. Delimitations and Limitations

A delimitation:

This study is confined to assessing and observing U.S. Army Soldiers assigned to two companies within a combat support training battalion at Fort Leonard Wood, Missouri.

Limitations:

The study sample consists of Soldier accessions into military occupational specialties within the U.S. Army's combat support branch. Since combat support units may differ from combat arms and combat service support units in terms of the distributions of sex and personnel aptitudes, this study may not be generalizable to all Army training programs.

The study sample consists of Soldier accessions into the U.S. Army in the month of August. Since the demographics of Soldiers entering Basic Combat Training exhibit a seasonal variation, the findings of this study may not directly apply to other Basic Combat Training classes at the study location.

C. METHODS

1. Research Design

The study protocol was approved by the Naval Postgraduate School Institutional Review Board in accordance with 32 Code of Federal Regulations 219 and SECNAV Instruction 3900.39D. The study used a quasi-experimental study design that was embedded within the Army's 63-day Basic Combat Training program of instruction. The intervention and comparison groups were selected without random assignment, although group assignment to the treatment condition was random. Participant assignment to group was made by the U.S. Army based on factors that were unobservable by the research team, but which were not altered for the purpose of this study. That is, the research team took the groups as they were created by the U.S. Army based on their normal mode of operations for managing Basic Combat Training. The study intervention consisted of a modification of the timing of sleep and wake periods; otherwise, no change was made to the content, instructional methods, or sequence of Basic Combat Training

events. The intervention group used a phase-delayed (i.e., 11:00 p.m. – 7:00 a.m.) sleep regimen with opportune midday naps, while the comparison group maintained the standard (i.e. 8:30 p.m. – 4:30 a.m.) sleep regimen. The barracks used by the intervention group were modified with black-out curtains to mitigate the effect of morning light; no modifications were made to the barracks used by the comparison group.

2. Participants

Participants for the comparison group were solicited from among those Soldiers starting Basic Combat Training on August 14, 2009, and assigned to Charlie Company, 3rd Battalion, 10th Infantry Regiment, 3rd Chemical Brigade (C/3-10 IN BN), Fort Leonard Wood, Missouri. Similarly, participants for the intervention group were solicited from among those Soldiers starting Basic Combat Training on August 21, 2009, and assigned to Bravo Company, 3rd Battalion, 10th Infantry Regiment (B/3-10 IN BN). Participants for both groups were solicited during Basic Combat Training in-processing by a civilian member of the research team to mitigate the potential for implied coercion by rank. Soldiers who chose not to participant in the study (less than 1%) still followed the training company's schedule and accomplished all training events, but they did not complete any of the study-related instruments.

3. Data Collection Instruments and Variables

a. Actiwatch

The Actiwatch® (Model AW-64, Philips Respironics, Bend, Oregon) is a 16-gram, 28 x 27 x 10-millimeter wristwatch-like device worn on the nondominant wrist that objectively measures activity and rest patterns. With each participant movement, a highly sensitive accelerometer generates a variable voltage that is digitally processed and sampled at a frequency of 32 Hertz. The signal is integrated over a user-selected epoch and a value expressed as activity counts is recorded in the on-board memory. Data are downloaded to a computer and may be expressed graphically as an actogram or reported in American standard code for information interchange (ASCII) format numerically as total activity counts per epoch.

b. Basic Rifle Marksmanship

Objective evaluation of rifle marksmanship skill was made based on "record fire" score. During a Basic Combat Training record fire, Soldiers are given an M16/M4 series rifle and 40 rounds of ammunition and presented with 40 timed target exposures at ranges from 50 to 300 meters. Twenty targets are engaged with 20 rounds from the prone supported position, ten targets are engaged with ten rounds from the prone unsupported position, and ten targets are engaged with ten rounds from the kneeling position—while wearing a helmet and load-bearing equipment. The standard is to obtain at least 23 target hits on the 40 targets exposed. Soldiers complete a practice record fire on days 29 and 30 of Basic Combat Training and an official record fire on day 32 of Basic Combat Training, for a total of three sequential record fires (Directorate Basic Combat Training Doctrine and Training Development, 2008, March).

c. General Technical Aptitude

Objective evaluation of individual aptitude was made based on General Technical (GT) score as derived from the Armed Services Vocational Aptitude Battery (ASVAB). The ASVAB is a 216-item inventory containing nine separately timed subtests: General Science, Arithmetic Reasoning, Word Knowledge, Paragraph Comprehension, Auto and Shop, Mathematics Knowledge, Mechanical Comprehension, Electronics Information, and Assembling Objects. The ASVAB is not an intelligence test, but rather, is specifically designed to measure an individual's aptitude to be trained in specific jobs. GT score is a composite of the Arithmetic Reasoning, Word Knowledge, and Paragraph Comprehension subtests, and it is often a major determinant of the occupational specialties for which a person can be considered in the military.

d. Mood State

Subjective evaluation of mood was made with the Profile of Mood States (POMS) (McNair, Lorr, & Droppleman, 1981). The POMS is a 65-item questionnaire that measures affect or mood on 6 scales: 1) tension-anxiety, 2) depression-dejection, 3) anger-hostility, 4) vigor-activity, 5) fatigue-inertia, and 6) confusion-bewilderment. An

aggregate mood disturbance score is calculated by summing the scores on the six scales and negatively weighting the vigor-activity score.

e. Personality

A personality assessment was accomplished using the Neuroticism-Extroversion-Openness Five-Factor Inventory (NEO-FFI) (Costa & McCrae, 1992). The NEO-FFI is essentially a short form of the Revised NEO Personality Inventory (NEO-PI-R). It consists of 60 items from the NEO-PI-R that are used to score the five domains: 1) neuroticism, 2) extraversion, 3) openness, 4) agreeableness, and 5) conscientiousness. It does not contain the items for assessing the facets within each domain. The NEO-FFI is designed for use in circumstances in which time is too limited to present the entire NEO-PI-R or only scores on the five domains are required (Weiner & Greene, 2008).

f. Physical Fitness

Objective evaluation of physical fitness was made based on Army Physical Fitness Test (APFT) score. Soldiers complete a physical fitness assessment consisting of three measured events: push-ups, sit-ups, and a timed 2-mile run. Raw scores are scaled for both age and sex. Soldiers must earn a score of 150 points or higher on the end-of-training APFT with 50 points or more in each event to graduate from Basic Combat Training (Directorate Basic Combat Training Doctrine and Training Development, 2008). Soldiers complete two diagnostic APFTs during the third and sixth weeks of Basic Combat Training and a final APFT in the eighth week of training.

g. Resilience

Assessment of resilience to stress was accomplished using the Response to Stressful Experiences Scale (RSES) (Johnson et al., 2008). The RSES was developed by researchers with the National Center for Post Traumatic Stress Disorder to rate psychological traits that promote resilience, which is the ability to undergo stress and still retain mental health and well-being. It consists of 22 items and identifies six factors that are key to psychological resilience: 1) positive outlook, 2) spirituality, 3) active coping, 4) self-confidence, 5) learning and making meaning, and 6) acceptance of limits. The

RSES has been tested on more than 1,000 active-duty military personnel (Naval Center for Combat and Operational Stress Control, 2009).

h. Sleep Habits

Subjective assessments of sleep habits were made using three validated survey instruments. The first instrument was the Pittsburgh Sleep Quality Index (PSQI), a self-rated questionnaire designed to measure sleep quality in clinical populations by looking at sleep in the previous month. Nineteen individual items generate the following seven scores: 1) subjective sleep quality, 2) sleep latency, 3) sleep duration, 4) habitual sleep efficiency, 5) sleep disturbances, 6) use of sleeping medications, and 7) daytime dysfunction. A review of this survey's reliability asserts that the PSQI is useful to both psychiatric clinical practice and research activities (Buysse, Reynolds, Monk, Berman, and Kupfer, 1989).

The second instrument was the Epworth Sleepiness Scale (ESS) (Johns, 1991), an 8-item scale commonly used to diagnose sleep disorders and considered a valid and reliable self-report of sleepiness. Participants use an integer number from 0 to 3, corresponding to the likelihood (never, slight, moderate, and high, respectively) that they would fall asleep in eight situations such as sitting and reading, watching television, as a passenger in a car for an hour, etc. Ratings above 10 out of a possible 24 are cause for concern with respect to an underlying sleep disorder.

The third instrument was the Morningness-Eveningness Questionnaire (MEQ) published by Horne and Ostberg (1976), which contains 19 questions aimed at determining when, during the daily temporal span, individuals have the maximum propensity to be active. Most questions are preferential, in the sense that the respondent is asked to indicate when they would prefer, rather than when they actually do, wake up or begin sleep. Questions are multiple-choice and each answer is assigned a value such that their sum gives a score ranging from 16 to 86, with lower values corresponding to evening chronotypes and higher values indicating morning chronotypes.

i. Study Questionnaires

The pre-study questionnaire contained ten questions aimed at potential covariates that could influence study outcome measures. Four questions asked participants for their age, sex, height, and weight. One question asked participants to quantify their frequency of exercise during the preceding month, both in terms of the number and duration of exercise sessions. Another question asked whether participants regularly used firearm(s), and if so, to characterize the type of firearm(s), reason(s) for use, and frequency of use. Three questions addressed use of caffeinated beverages, tobacco, and medications. Lastly, one question asked participants to quantify the amount of sleep per day they required to feel ready to start the day.

The post-study questionnaire consisted of six questions. Similar to the pre-test questionnaire, two questions addressed use of caffeinated beverages and medications, and one question asked participants to quantify the amount of sleep per day they required to feel ready to start the day. One question asked participants about the frequency with which they fell asleep during activities. Another question asked participants to provide an ordinal ranking on a 5-item Likert scale of the adequacy of both their sleep and that of their peers during Basic Combat Training. The final question asked participants' preference for the timing of daily physical training.

4. Procedures

a. General

Prior to beginning the study, each participant received a full briefing on the purposes of the study and assurances about the confidentiality of the data. Once informed consent was obtained, each participant completed the pre-study questionnaire followed by the Epworth Sleepiness Scale (ESS), Pittsburgh Sleep Quality Index (PSQI), Morningness-Eveningness Questionnaire (MEQ), Response to Stressful Experiences Scale (RSES), Profile of Mood States (POMS), and Neuroticism-Extroversion-Openness (NEO) Five Factor Inventory (Table VI-2). Participants subsequently accomplished the POMS at weekly intervals throughout Basic Combat Training. At the completion of Basic Combat Training, participants received an out-briefing and completed the post-

study questionnaire followed by the ESS, PSQI, and the final POMS. For each participant, data were collected on general technical aptitude, basic rifle marksmanship, and physical fitness scores from preexisting local databases. Attritions were determined from training company graduation rosters.

Table VI-2. Schedule for data-generating events.

↓Data Event Week→	1	2	3	4	5	6	7	8	9
Actigraphy*	X	X	X	X	X	X	X	X	X
Army Physical Fitness Test			X			X		X	
Basic Rifle Marksmanship					X				
Epworth Sleepiness Scale	X								X
Morningness-Eveningness Questionnaire	X								
NEO Five-Factor Inventory	X								
Pittsburgh Sleep Quality Index	X								X
Profile of Mood States	X	X	X	X	X	X	X	X	X
Response to Stressful Experiences Scale	X								
Study Questionnaires	X								X

*Actigraphy data was collected on a random subsample of the study participants.

b. Actigraphy

A random sample comprising approximately 20% of participants in each study group was selected for actigraphic data collection. Participants agreeing to actigraphic data collection were issued an Actiwatch® on Day 1 to track sleep and activity patterns in a relatively unobtrusive fashion. Participants were asked to wear the Actiwatch® continuously on the wrist of their nondominant hand during all waking and sleeping periods and not to remove it for showering. The Actiwatch® was collected from each participant during Week 4 (intervention group) or Week 5 (comparison group) for downloading of data and reinitialization of the Actiwatch® data collection mode. Once

the data collection period was complete, the data were taken back to the laboratory and, using Actiware® version 5.57.0006 software, scored for sleep times.

c. Statistical Analysis

For the pre-study and post-study questionnaires and the ESS, PSQI, MEQ, and RSES survey instruments, item nonresponse was handled using stochastic regression imputation to reduce the bias that could be caused by ignoring records with missing data (Kim & Curry, 1977; Brick & Kalton, 1996). For the NEO-FFI and the POMS survey instruments, item nonresponse was handled per the guidance in the associated survey technical manuals. In the case of the weekly POMS, which were administered repetitively throughout the course of training, no attempt was made to address unit or partial nonresponses. Microsoft® Office Excel® 2007 was used to develop the study database; histograms of the actigraphy data were created using the Analysis ToolPak add-in. Analyses were undertaken with the Statistical Package for the Social Sciences (SPSS) version 11. All data were assessed for normalcy, and parametric and nonparametric approaches were used accordingly for descriptive statistical analyses. Separate univariate and repeated measures analyses of covariance (ANCOVAs) were used to test major hypotheses involving measures with one dependent variable. Repeated measures were analyzed using a univariate approach with a fixed effect for time when there were a substantial number of unit nonresponses, thereby reducing the danger of biased repeated measures estimates of treatment effects caused by ignoring records with missing responses. ANCOVA results were examined to determine whether there were sphericity violations of sufficient magnitude to warrant the use of Huynh-Feldt adjusted degrees of freedom. Multivariate analysis of covariance (MANCOVA) was used to test hypotheses involving measures with more than one dependent variable. Box's and Levene's tests were used to assure the multivariate assumptions of equality of covariance matrices and that equality of error variances across groups was not violated. Lastly, logistic regression was used to test major hypotheses involving measures with a binary dependent variable.

D. RESULTS

1. Participants (*n* = 392)

The study sample was comprised of 392 participants, 209 in the intervention group and 183 in the comparison group. Participants' responses on the pre-study questionnaire and survey instruments are summarized in Tables VI-3 through VI-5 by treatment condition, that being either assignment to the intervention or comparison group. Figures VI-4 through VI-6 display histograms for a select subset of questions from the PSQI asking participants about their baseline sleep schedule. From the outset of the study, the intervention and comparison groups were generally comparable, although they did differ on some of the measured variables:

1) Participants in the intervention group tended to have a higher body mass index (i.e., body weight corrected for height) than those in the comparison group.
2) A greater proportion of participants in the intervention group were in the National Guard/Reserves as compared to the comparison group.
3) Participants in the comparison group reported higher levels of neuroticism on the NEO-FFI, while participants in the intervention group reported higher levels of conscientiousness.
4) Participants in the comparison group tended to have higher global scores on the pre-study PSQI, mainly because of increased daytime dysfunction. Also, a greater proportion of participants in the comparison group met the threshold score for being classified as potentially having poor quality sleep.
5) Participants in the intervention group had higher levels of spirituality, active coping, and self-efficacy, and hence, overall resilience, as assessed by the RSES at the outset of the study.

Table VI-3. Summary of intervention and comparison study groups at outset of study.

Variable	Group: Intervention (n = 209)	Group: Comparison (n = 183)	p-value
Age (yrs), median (IQR)	20 (18-23)	20 (18-24)	0.762^{M}
Body mass index ($kg \cdot m^{-2}$), median (IQR)	25.4 (22.9-28.4)	23.6 (21.6-26.8)	$0.002^{M}*$
Body mass index category, no. (%)			
Underweight	5 (2.4)	6 (3.3)	
Normal	87 (41.6)	102 (55.7)	$0.021^{C}*$
Overweight	81 (38.8)	57 (31.1)	
Obese	36 (17.2)	18 (9.8)	
Caffeine			
Consume caffeinated beverages, no. (%)	116 (55.5)	110 (60.1)	0.357^{C}
Caffeine use ($mg \cdot d^{-1}$), median (IQR)	39.0 (0-157.5)	61.0 (0-177.0)	0.248^{M}
Component, no. (%)			
National Guard	72 (34.4)	58 (31.7)	
Regular	82 (39.2)	109 (59.6)	$< 0.001^{C}*$
Reserves	55 (26.3)	16 (8.7)	
Epworth Sleepiness Scale			
Total score, median (IQR)	8 (6-11)	9 (6-11)	0.562^{M}
Excessive fatigue (score > 10), no. (%)	52 (24.9)	52 (28.4)	0.429^{C}
Exercise frequency ($hrs \cdot wk^{-1}$), median (IQR)	2.5 (1.0-4.5)	3.0 (1.5-5.9)	0.071^{M}
Firearms			
Regularly use firearm, no. (%)	51 (24.4)	39 (21.3)	0.468^{C}
Type of firearm, no. (%)			
Rifle	44 (21.1)	31(16.9)	0.302^{C}
Handgun	28 (13.4)	23 (12.6)	0.808^{C}
Use of firearm, no. (%)			
Hunting	36 (17.2)	28 (15.3)	0.607^{C}
Sport shooting	32 (15.3)	28(15.3)	0.998^{C}
Other	7 (3.8)	4 (1.9)	0.253^{C}
Frequency of use ($days \cdot yr^{-1}$), median (IQR)	0 (0-0)	0 (0-0)	0.540^{M}

*Significant at ≤ 0.05 level.
^{C}Chi square statistic, ^{M}Mann-Whitney U.
Note: IQR = interquartile range.

Table VI-4. Summary of intervention and comparison study groups at outset of study (continued).

Variable	Group: Intervention (n = 209)	Group: Comparison (n = 183)	p-value
GT score, median (IQR)	105 (96-114)	108 (99-116)	0.057^{M}
Morningness-Eveningness Questionnaire			
Total score, median (IQR)	50 (45-55)	49 (42-56)	0.498^{M}
Chronotype, no (%)			
Evening type	39 (18.7)	34 (18.6)	0.291^{C}
Neither type	140 (67.0)	112 (61.2)	
Morning type	30 (14.3)	37 (20.2)	
NEO Five Factor Inventory, median (IQR)			
Neuroticism	52 (45-59)	55 (47-63)	0.012^{M}*
Extraversion	53 (46-61)	53 (46-60)	0.601^{M}
Openness to experience	48 (41-58)	50 (41-57)	0.712^{M}
Agreeableness	46 (36-53)	44 (36-52)	0.224^{M}
Conscientiousness	50 (43-57)	46 (38-53)	0.003^{M}*
Pittsburgh Sleep Quality Index			
Global score, median (IQR)	6 (4-9)	7 (5-10)	0.048^{M}*
Poor sleep quality (score > 5), no. (%)	123 (58.9%)	129 (70.5%)	0.016^{C}*
Component scores, median (IQR)			
Subjective sleep quality	1 (1-1)	1 (1-2)	0.190^{M}
Sleep latency	2 (1-4)	2 (1-4)	0.817^{M}
Sleep duration	0 (0-2)	0 (0-2)	0.430^{M}
Habitual sleep efficiency	0 (0-0)	0 (0-0)	0.203^{M}
Sleep disturbances	1 (1-2)	1 (1-2)	0.399^{M}
Use of sleeping medication	0 (0-0)	0 (0-0)	0.400^{M}
Daytime dysfunction	1 (0-1)	1 (1-1)	0.001^{M}*
Rank, no (%)			
E01	82 (39.2)	62 (33.9)	0.514^{C}
E02	69 (33.0)	58 (31.7)	
E03	43 (20.6)	48 (26.2)	
E04	15 (7.2)	15 (8.2)	

*Significant at ≤ 0.05 level.
^{C}Chi square statistic, ^{M}Mann-Whitney U.
Note: IQR = interquartile range.

Table VI-5. Summary of intervention and comparison study groups at outset of study (continued).

Variable	Group: Intervention (n = 209)	Group: Comparison (n = 183)	p-value
Response to Stressful Experiences Scale			
Global score, median (IQR)	69 (60-78)	67 (58-75)	0.008^{M}*
Factor scores, median (IQR)			
Positive appraisal	7.3 (6.2-8.3)	7.0 (5.9-8.0)	0.141^{M}
Spirituality	2.9 (2.9-3.8)	2.9 (2.7-3.8)	0.001^{M}*
Active coping	10.8 (8.9-12.2)	10.2 (8.2-11.5)	0.001^{M}*
Self-efficacy	3.2 (2.4-3.2)	2.4 (2.4-3.2)	0.029^{M}*
Learning and meaning-making	6.6 (5.4-8.0)	6.5 (5.1-7.3)	0.025^{M}*
Acceptance of limitations	4.9 (3.5-5.6)	4.3 (3.5-5.0)	0.055^{M}*
Sex, no. (%)			
Female	67 (32.1)	52 (28.4)	0.434^{C}
Male	142 (67.9)	131 (71.6)	
Tobacco			
Regularly use tobacco, no (%)	81 (38.8)	68 (37.2)	0.745^{C}
Frequency of use ($cigs \cdot wk^{-1}$), median (IQR)	0 (0-28)	0 (0-16)	0.519^{M}

*Significant at ≤ 0.05 level.
^{C}Chi square statistic, ^{M}Mann-Whitney U.
Note: IQR = interquartile range.

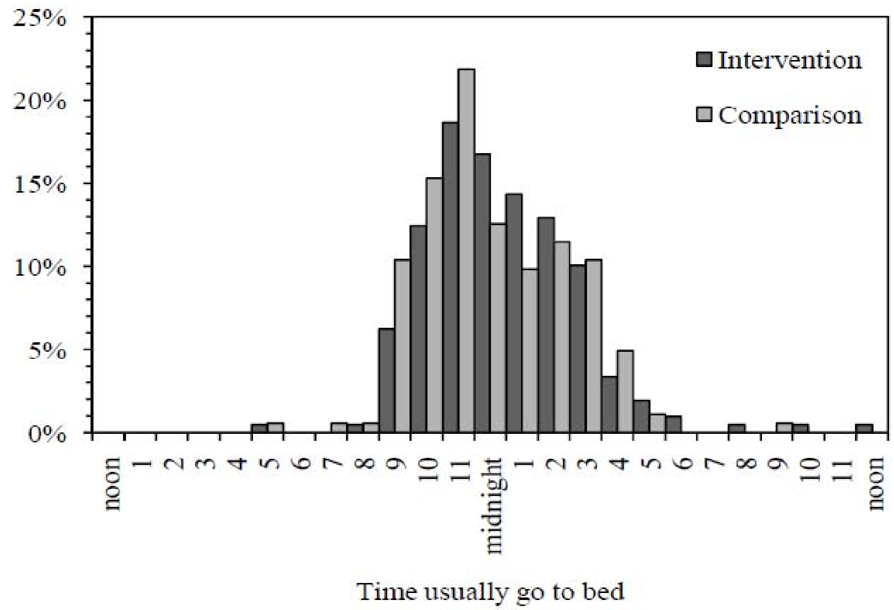


Figure VI-4. Histogram of participants' reported usual bed time (PSQI question 1).

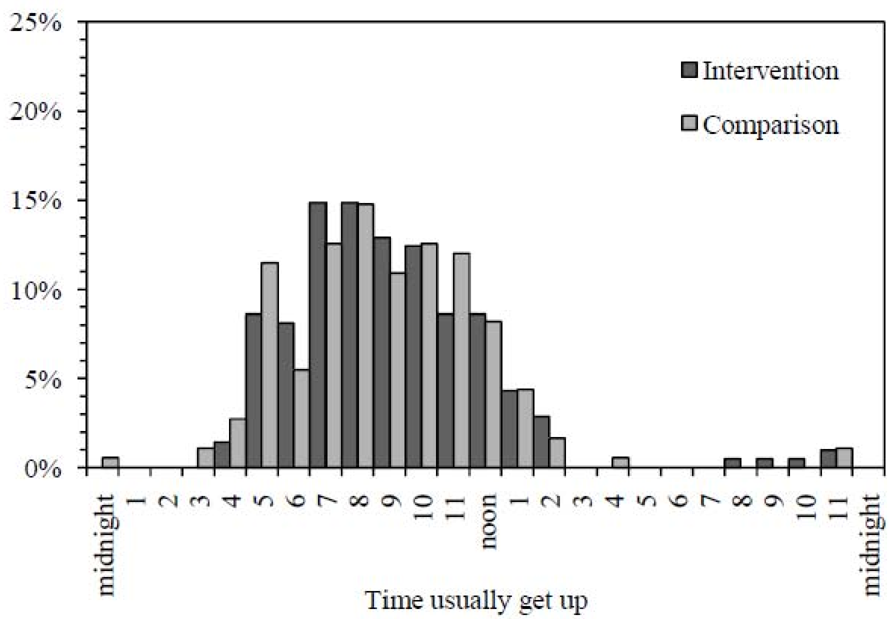


Figure VI-5. Histogram of participants' reported usual getting up time (PSQI question 3).

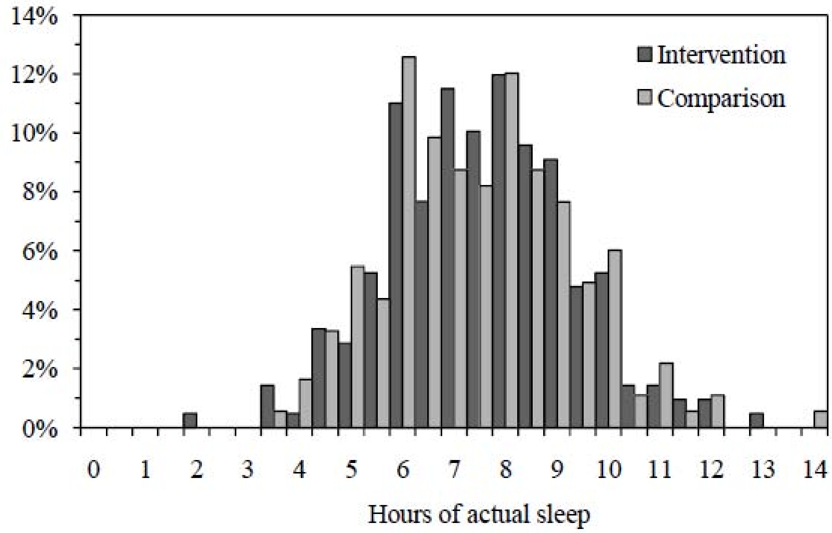


Figure VI-6. Histogram of participants' reported hours of sleep per night (PSQI question 4).

2. Actigraphy Subsample

a. Participants (*n* = 95)

What follows in this subsection is limited to the subsample of 95 participants, 53 in the intervention group and 42 in the comparison group, randomly selected to wear Actiwatches®. Due to unexplained technical difficulties, data were not recorded on Actiwatches® given to one participant in the comparison group. Consequently, this participant's other data were censored in the subsequent analysis, thereby leaving us with a subsample of 94 participants. Across the subsample, on average, 83.8 (standard deviation 9.6; range 36-92) participants had a valid Actiware® score for any given day of Basic Combat Training. A one-way analysis of variance (ANOVA) was used to compare the number of participants per day with a valid Actiware® score by week of training. Overall, there was a significant difference in week ($F_{8,52}$ = 3.205, p = 0.005), but Bonferroni post-hoc tests showed that this difference was only between Week 2 (mean 90.7 participants) and Week 9 (mean 73.4 participants).

Participants' responses on the study questionnaire and survey instruments are summarized in Tables VI-6 through VI-8 by treatment condition, that being either

assignment to the intervention or comparison group. Figures VI-7 through VI-9 display histograms for a select subset of questions from the PSQI asking participants about their baseline sleep schedule. From the outset of the study, the intervention and comparison groups were comparable on practically all measured variables. The only statistically significant difference between groups was the percentage of those handling firearms who reported using a rifle. All the participants in the intervention group who reported handling firearms used a rifle, while slightly more than half of those in the comparison group did so. There was also a tendency for participants in the intervention group to have a higher body mass index than those in the comparison group, but this difference was not statistically significant. Likewise, there was a tendency for a greater proportion of participants in the intervention group to be in the National Guard/Reserves as compared to the comparison group, but this difference was also not statistically significant.

Table VI-6. Summary of intervention and comparison study groups for actigraphy subsample at outset of study.

Variable	Group: Intervention (n = 53, 25%)	Group: Comparison (n = 41, 22%)	p-value
Age (yrs), median (IQR)	19 (18-23)	20 (18-24)	0.320^{M}
Body mass index ($kg \cdot m^{-2}$), median (IQR)	25.1 (22.2-27.8)	23.1 (21.4-26.0)	0.074^{M}
Body mass index category, no. (%)			
Underweight	1 (1.9)	1 (2.4)	0.232^{V}
Normal	24 (45.3)	27 (65.9)	
Overweight	18 (34.0)	9 (22.0)	
Obese	10 (18.9)	4 (9.8)	
Caffeine			
Consume caffeinated beverages, no. (%)	35 (66.0)	20 (48.8)	0.092^{C}
Caffeine use ($mg \cdot d^{-1}$), median (IQR)	164 (108-288)	144 (72-305)	0.327^{M}
Component, no. (%)			
National Guard/Reserve	30 (56.6)	16 (39.0)	0.091^{C}
Regular	23 (43.4)	25 (61.0)	
Epworth Sleepiness Scale			
Total score, mean (SD)	7.9 (3.2)	7.4 (3.5)	0.473^{T}
Excessive fatigue (score > 10), no. (%)	9 (17.0)	7 (17.1)	0.991^{C}
Exercise frequency ($hrs \cdot wk^{-1}$), median (IQR)	2.0 (1.4-4.2)	3.0 (1.4-6)	0.226^{M}
Firearms			
Regularly use firearm, no. (%)	11 (20.8)	7 (17.1)	0.653^{C}
Type of firearm, no. (%)			
Rifle	11 (100)	4 (57.1)	0.043^{F}*
Handgun	4 (36.4)	4 (57.1)	0.630^{F}
Use of firearm, no. (%)			
Hunting	7 (63.6)	3 (42.9)	0.630^{F}
Sport shooting	8 (72.7)	4 (57.1)	0.627^{F}
Other	0 (0)	2(28.6)	0.137^{F}
Frequency of use ($days \cdot yr^{-1}$), median (IQR)	30 (20-45)	45 (25-50)	0.340^{M}

*Significant at ≤ 0.05 level.
^{C}Chi square statistic, ^{F}Fisher's Exact Test, ^{M}Mann-Whitney U, ^{T}Student's t-test, ^{V}Cramer's V.
Notes: IQR = interquartile range; SD = standard deviation.

Table VI-7. Summary of intervention and comparison study groups for actigraphy subsample at outset of study (continued).

Variable	Group: Intervention (n = 53, 25%)	Group: Comparison (n = 41, 22%)	p-value
GT score, median (IQR)	108 (96-116)	110 (99-121)	0.354[M]
Morningness-Eveningness Questionnaire			
Total score, mean (SD)	50.6 (8.9)	47.2 (9.7)	0.086[T]
Chronotype, no (%)			
Evening type	11 (20.8)	15 (36.6)	0.226[C]
Neither type	31 (58.5)	20 (48.8)	
Morning type	11 (20.8)	6 (14.6)	
NEO Five Factor Inventory			
Neuroticism, median (IQR)	52 (44-56)	51 (46-63)	0.706[M]
Extraversion, mean (SD)	53.5 (11.5)	54.1 (9.0)	0.786[T]
Openness to experience, mean (SD)	50.7 (12.6)	49.7 (11.1)	0.683[T]
Agreeableness, mean (SD)	45.4 (11.4)	43.7 (11.4)	0.495[T]
Conscientiousness, median (IQR)	46 (42-59)	48 (41-53)	0.359[M]
Pittsburgh Sleep Quality Index			
Global score, mean (SD)	6.3 (2.5)	6.71 (2.8)	0.468[T]
Poor sleep quality (score > 5), no. (%)	32 (60.4)	28 (68.3)	0.428[C]
Component scores, median (IQR)			
Subjective sleep quality	1 (1-1)	1 (1-2)	0.147[M]
Sleep latency	2 (1-4)	2 (1-3)	0.745[M]
Sleep duration	0 (0-1)	0 (0-1)	0.504[M]
Habitual sleep efficiency	0 (0-0)	0 (0-0)	0.211[M]
Sleep disturbances	1 (1-2)	1 (1-2)	0.114[M]
Use of sleeping medication	0 (0-0)	0 (0-0)	0.699[M]
Daytime dysfunction	1 (0-1)	1 (0-1)	0.378[M]
Rank, no (%)			
E01	18 (34.0)	16 (39.0)	0.759[C]
E02	20 (37.7)	12 (29.3)	
E03	12 (22.6)	9 (22.0)	
E04	3 (5.7)	4 (9.8)	

[C]Chi square statistic, [M]Mann-Whitney U, [T]Student's *t*-test.
Notes: IQR = interquartile range; SD = standard deviation.

Table VI-8. Summary of intervention and comparison study groups for actigraphy subsample at outset of study (continued).

Variable	Group		*p*-value
	Intervention (n = 53, 25%)	Comparison (n = 41, 22%)	
Response to Stressful Experiences Scale			
Global score, mean (SD)	68.3 (12.0)	65.1 (13.0)	0.233^{T}
Factor scores, median (IQR)			
Positive appraisal	7.6 (6.1-8.3)	6.8 (6.2-8.5)	0.819^{M}
Spirituality	2.9 (2.9-3.8)	2.9 (2.9-3.8)	0.716^{M}
Active coping	8.7 (10.2-11.9)	10.2 (8.4-11.5)	0.778^{M}
Self-efficacy	3.2 (2.4-3.2)	2.4 (2.4-3.2)	0.778^{M}
Learning and meaning-making	7.2 (5.0-8.0)	6.5 (5.4-8.3)	0.310^{M}
Acceptance of limitations	4.3 (3.5-5.6)	4.3 (3.5-5.6)	0.816^{M}
Sex, no. (%)			
Female	20 (37.7)	15 (36.6)	0.909^{C}
Male	33 (62.3)	26 (63.4)	
Tobacco			
Regularly use tobacco, no (%)	22 (41.5)	15 (36.6)	0.628^{C}
Frequency of use (cigs·wk^{-1}), median (IQR)	49 (19-101)	35 (8-105)	0.577^{M}

^{C}Chi square statistic, ^{M}Mann-Whitney U, ^{T}Student's *t*-test.
Notes: IQR = interquartile range; SD = standard deviation.

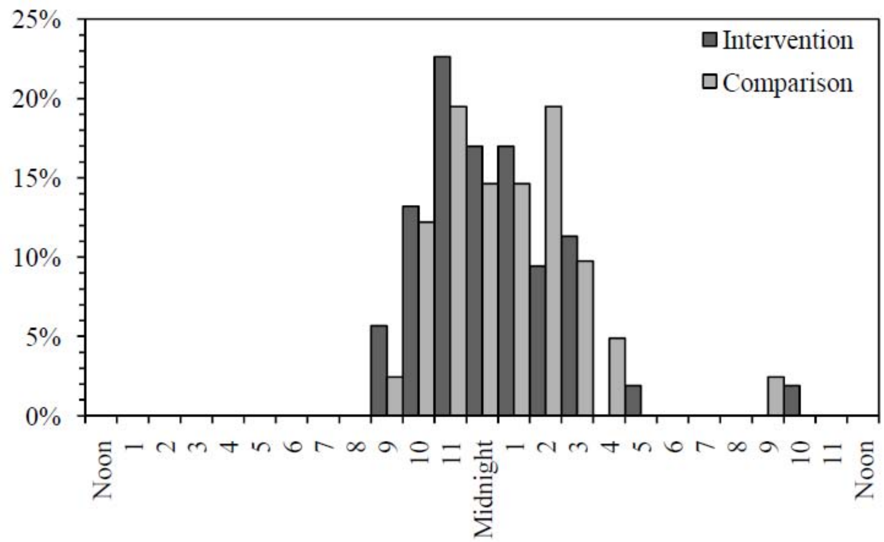


Figure VI-7. Histogram of participants' reported usual bed time (PSQI question 1) in actigraphy subsample.

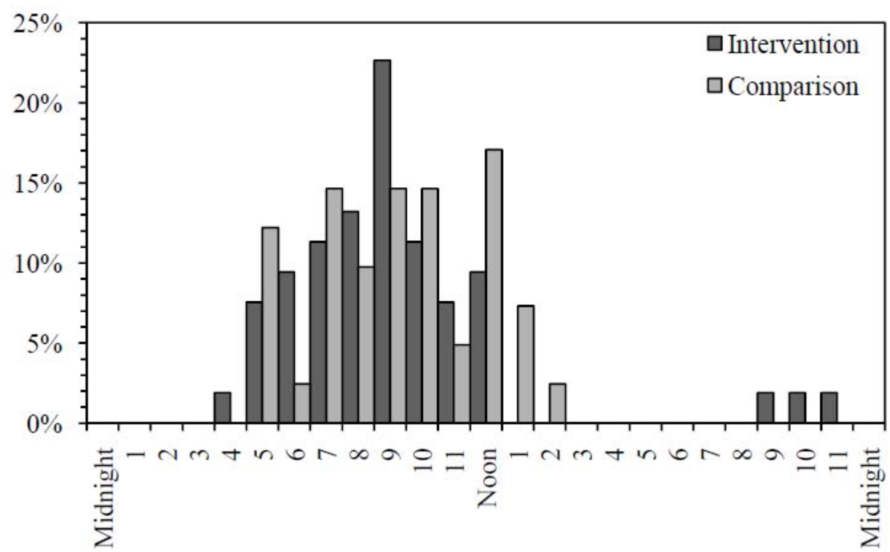


Figure VI-8. Histogram of participants' reported usual getting up time (PSQI question 3) in actigraphy subsample.

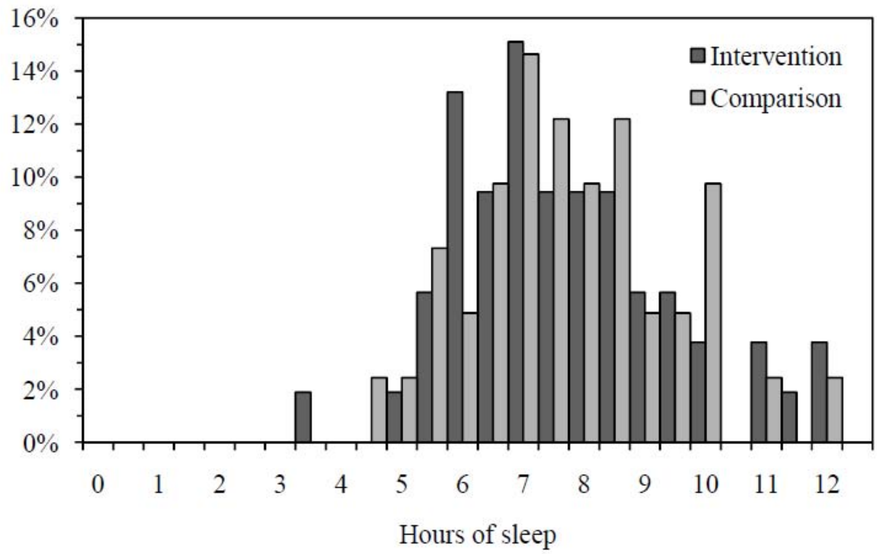
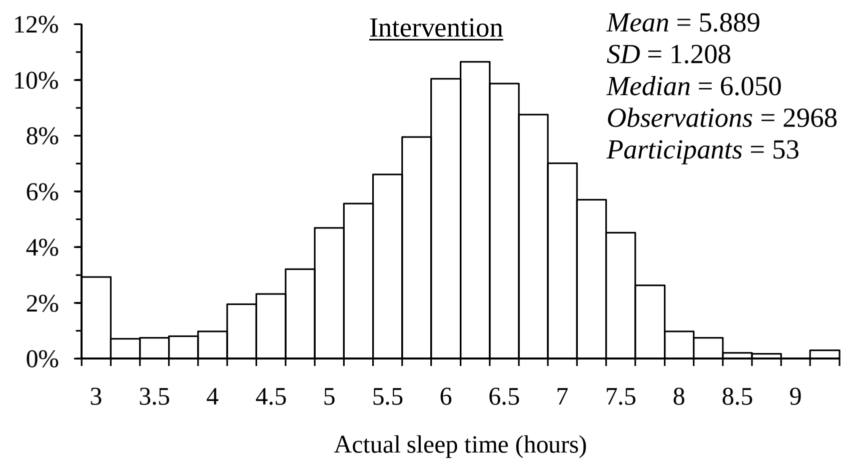


Figure VI-9. Histogram of participants' reported hours of sleep per night (PSQI question 4) in actigraphy subsample.

b. Total Sleep Time

Figure VI-10 shows the distribution and the parameters obtained from the distribution for daily total sleep obtained per night for all sleep observations gathered during Basic Combat Training according to treatment condition. The spike at 3 hours in both histograms was believed to be attributable to participants performing night watch duties. The median total sleep obtained per night across all weeks of Basic Combat Training was significantly greater for participants in the intervention versus comparison group (intervention group mean rank = 2,884.0; comparison group mean rank = 2,105.9; $p < 0.001$ based on Mann-Whitney U test). The National Sleep Foundation (NSF) recommends that adults obtain 7–9 hours of sleep per night. In this study, 15.5% of sleep observations in the intervention group satisfied the NSF recommendation versus only 4.6% in the comparison group—a significant difference ($\chi^2_1 = 152.282$, $p < 0.001$). Restated, the likelihood or odds of an episode of total daily sleep being less than the NSF's recommendation was 3.802 (95% CI: 3.037, 4.761) for the comparison group relative to the intervention group——i.e., they were nearly four times as likely to be sleep deficient in the comparison group.



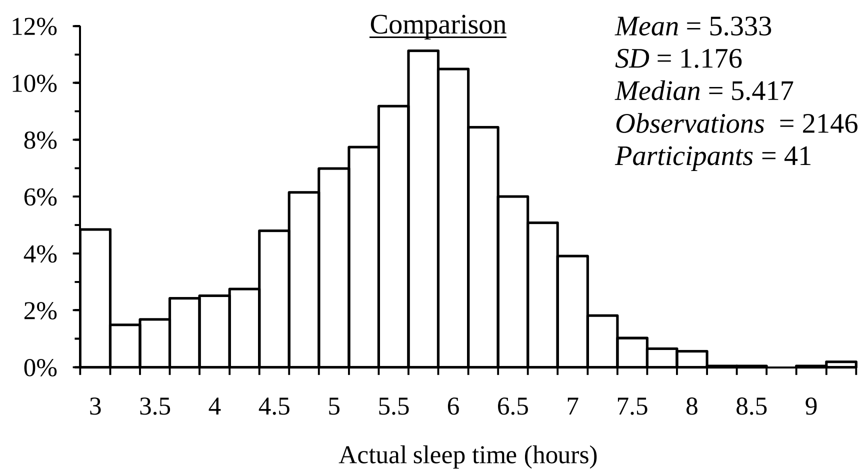


Figure VI-10. Histograms of total sleep obtained at night for all sleep observations gathered during Basic Combat Training according to treatment condition.

We examined how daily total sleep related to the treatment condition over the course of Basic Combat Training while accounting for potential covariates and the aforementioned differences between the study groups. However, any approach to analyzing total sleep time needed to address the issue that participants did not necessarily have valid Actiware® scores for every day of Basic Combat Training. This issue was remedied by first computing a weekly average sleep for each participant and then analyzing the dataset as a repeated cross-section design rather than a within-participant repeated measures design. A 1% significance level (or alpha of 0.01) was also used to

counter the resulting increased power of statistical tests. Accordingly, an ANCOVA of weekly average sleep was accomplished using treatment condition, week, and chronotype as fixed effects. Age, caffeine and tobacco use, component, firearm use, fitness factors (body mass index (BMI) and exercise frequency), GT score, personality component scores (NEO-FFI neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness scores), resilience (RSES score), sex, and sleep factors (ESS and PSQI scores) were covariates.

Table VI-9 provides the results of the univariate analysis of weekly average sleep. There was a significant fixed effect for treatment condition with an estimated marginal mean sleep for the intervention group of 5.876 (99% CI: 5.806, 5.945) versus 5.359 (99% CI: 5.276, 5.442) for the comparison group. That is, controlling for other variables, the intervention group obtained 31 minutes more sleep than the comparison group.

Table VI-9. Univariate tests for weekly average sleep.

Source	MS	df	*F*	*p*	η^2
Condition	32.384	1	140.162	<0.001*	0.163
Week	15.138	8	65.518	<0.001*	0.422
Chronotype	2.383	2	10.312	<0.001*	0.028
Condition x Week	2.555	8	11.059	<0.001*	0.110
Condition x Chronotype	0.323	2	1.399	0.247	0.004
Chronotype x Week	0.321	16	1.390	0.140	0.030
Condition x Chronotype x Week	0.116	16	0.502	0.947	0.011
Age	2.569	1	11.118	0.001*	0.015
Body mass index	1.476	1	6.390	0.012	0.009
Caffeine use (referent no)	2.490	1	10.779	0.001*	0.015
Component (referent regular)	0.232	1	1.004	0.317	0.001
Epworth Sleepiness Scale	2.491	1	10.781	0.001*	0.015
Exercise frequency	1.860	1	8.052	0.005*	0.011
Firearm use (referent no)	0.301	1	1.301	0.254	0.002
GT score	0.438	1	1.895	0.169	0.003
NEO-FFI					
Neuroticism	0.541	1	2.341	0.126	0.003
Extraversion	0.926	1	4.006	0.046	0.006
Openness to experience	0.090	1	0.387	0.534	0.001
Agreeableness	0.052	1	0.224	0.636	<0.001
Conscientiousness	0.937	1	4.055	0.044	0.006
Pittsburgh Sleep Quality Index	0.357	1	1.545	0.214	0.002
RSES	0.307	1	1.327	0.250	0.002
Sex (referent male)	2.376	1	10.285	0.001*	0.014
Tobacco use (referent no)	0.125	1	0.539	0.463	0.001
Error	0.231	718			

*Significant at ≤ 0.01 level.

Notes: GT score = General technical aptitude score; MS = Mean square; NEO-FFI = NEO Five-Factor Inventory; RSES = Response to Stressful Experiences Scale.

There was also a significant fixed effect for week (Figure VI-11), with pairwise differences occurring between week 1 versus weeks 6–9 ($p < 0.001$); week 2 versus weeks 6–9 ($p \leq 0.002$); week 3 versus week 6 and weeks 8–9 ($p < 0.001$); week 4 versus week 6 and weeks 8–9 ($p < 0.001$); week 5 versus weeks 6–9 ($p \leq 0.004$); week 6 versus week 7 ($p < 0.001$); and week 7 versus weeks 8–9 ($p < 0.001$).

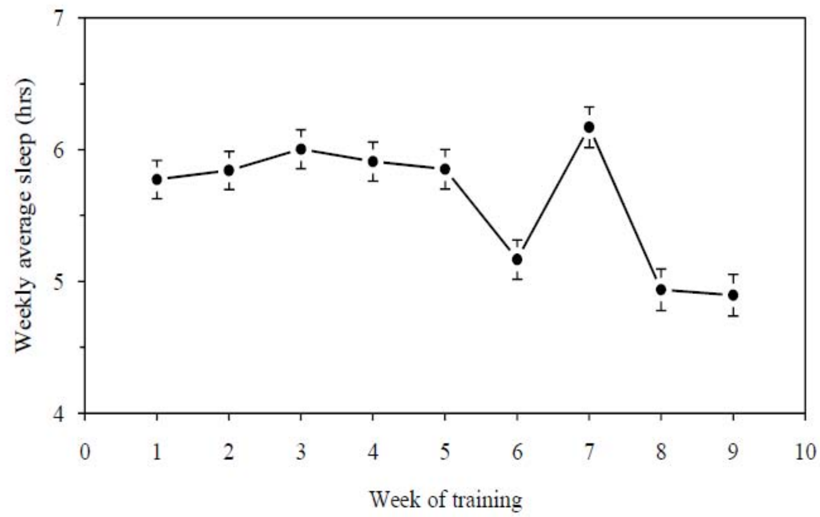


Figure VI-11. Estimated marginal means for sleep by week of training (error bars are for 99% confidence intervals).

For the significant fixed effect of chronotype (Figure VI-12), the pairwise differences occurred between morning chronotype versus both evening and indeterminate chronotypes ($p \leq 0.001$).

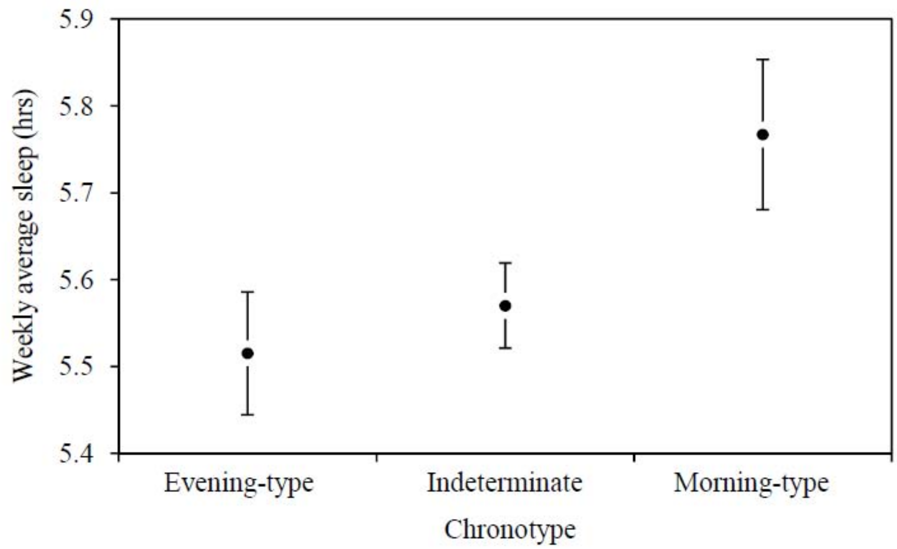


Figure VI-12. Estimated marginal means for sleep by chronotype (error bars are for 99% confidence intervals).

Additionally, there was a significant interaction effect between treatment condition and week (Figure VI-13), with participants in the intervention group getting more sleep than those in the comparison group during the first 6 weeks of training. During the latter three weeks of training, participants in the intervention group got notably less sleep such that there was no longer a difference between the intervention and comparison groups. This observation was attributed to the field exercises that were conducted throughout the last three weeks of training, during which participants moved from the barracks to an encampment. There was no interaction effect between treatment condition and chronotype or between chronotype and week. Significant covariates included age, caffeine use, ESS score, exercise frequency, and sex.

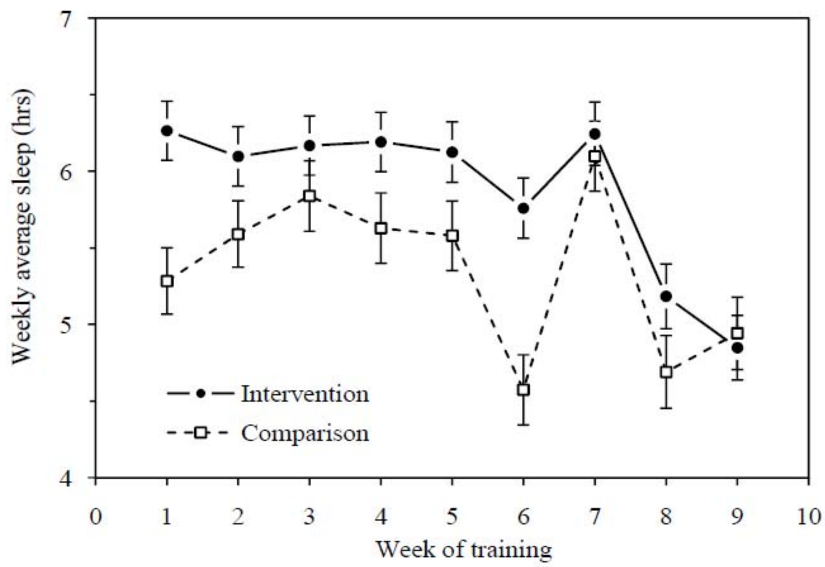


Figure VI-13. Estimated marginal means for sleep by treatment condition and week of training (error bars are for 99% confidence intervals).

c. Sleep Efficiency

Sleep efficiency was calculated as the ratio of a participant's total sleep time to total time in bed; it represents the proportion of time that a participant was assumed to be "in bed" or attempting sleep that was actually spent asleep (Paquet, Kawinska, & Carrier, 2007). Figure VI-14 shows the distribution and distributional parameters for sleep efficiency for all sleep observations gathered during Basic Combat Training according to treatment condition.

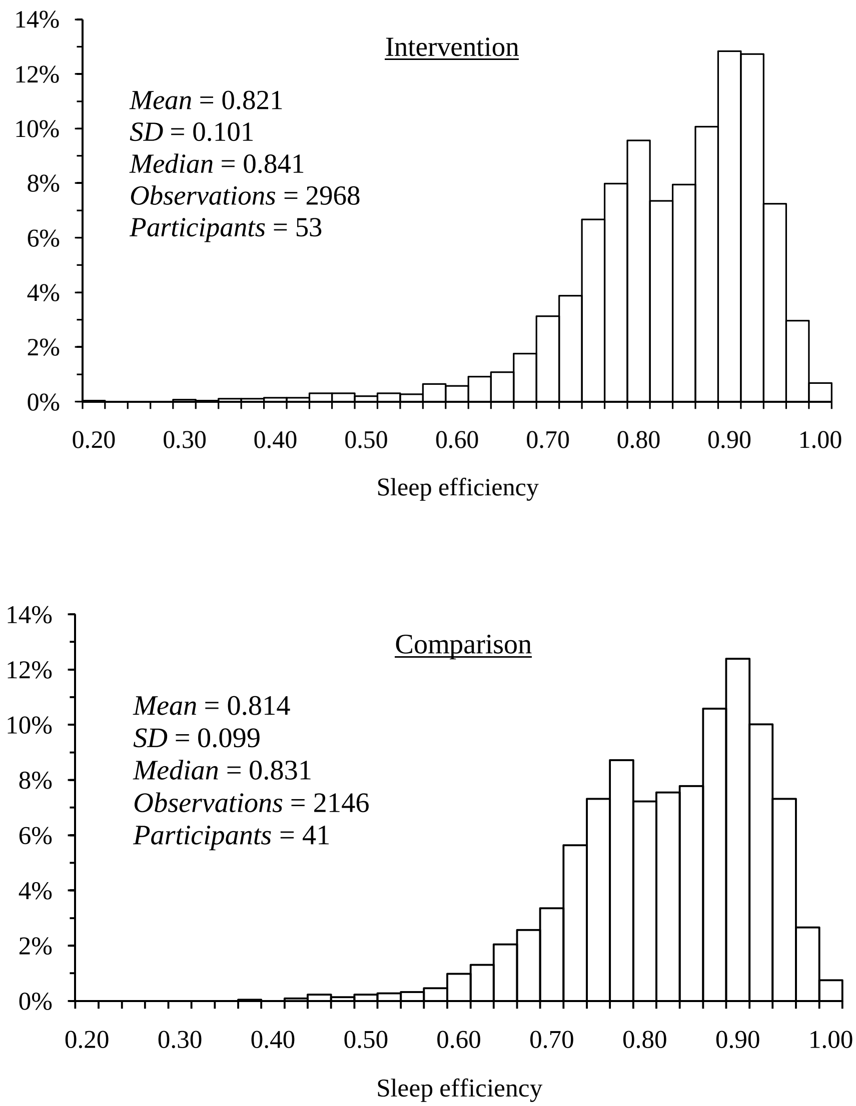


Figure VI-14. Histograms of sleep efficiency for all sleep observations by treatment condition.

The median sleep efficiency across all weeks of Basic Combat Training was significantly greater for participants in the intervention vice comparison study group (intervention group mean rank = 2,614.3; comparison group mean rank = 2,479.0; $p < 0.001$ based on Mann-Whitney test). Nevertheless, the practical significance of a difference in median sleep efficiency of 0.010 is questionable. However, the histograms suggest that the distributions of sleep efficiency for the two groups differed slightly. This impression was investigated further by estimating the population moments using the

sample k^{th} moments (Table VI-10). While the 95% confidence intervals overlapped for the first and second moments, there was a significant difference in the third and fourth moments, which are functions of the distributions' skewness (i.e., symmetry) and kurtosis (i.e., peakedness), respectively.

Table VI-10. Population moment estimates based on sample k^{th} moments.

	Intervention group		Comparison group	
k^{th} moment	Estimate	95% CI	Estimate	95% CI
First	0.821	(0.817, 0.825)	0.814	(0.810, 0.818)
Second	0.684	(0.678, 0.690)	0.672	(0.666, 0.679)
Third	0.577	(0.571, 0.584)	0.562	(0.555, 0.570)
Fourth	0.492	(0.485, 0.499)	0.476	(0.467, 0.484)

Note: CI = confidence interval.

d. Activity Counts During Sleep

Activity counts reflect movements during sleep and may be a function of the stage of sleep (Monk, Buysse, & Rose, 1999). Figure VI-15 shows the distribution and distributional parameters for mean activity counts for all sleep observations gathered during Basic Combat Training according to treatment condition. The median activity count during sleep across all weeks of Basic Combat Training was significantly less for participants in the intervention versus comparison study group (intervention group mean rank = 2,504.8; comparison group mean rank = 2,630.4; $p < 0.001$ based on Mann-Whitney test). However, the histograms appear quite similar; as in the analysis of the sleep efficiency data, population moments were estimated for each distribution using the k^{th} sample moments. It was found that the 95% confidence intervals overlapped for the first four moments of each sample distribution, thereby suggesting that the observed distributions do not significantly differ.

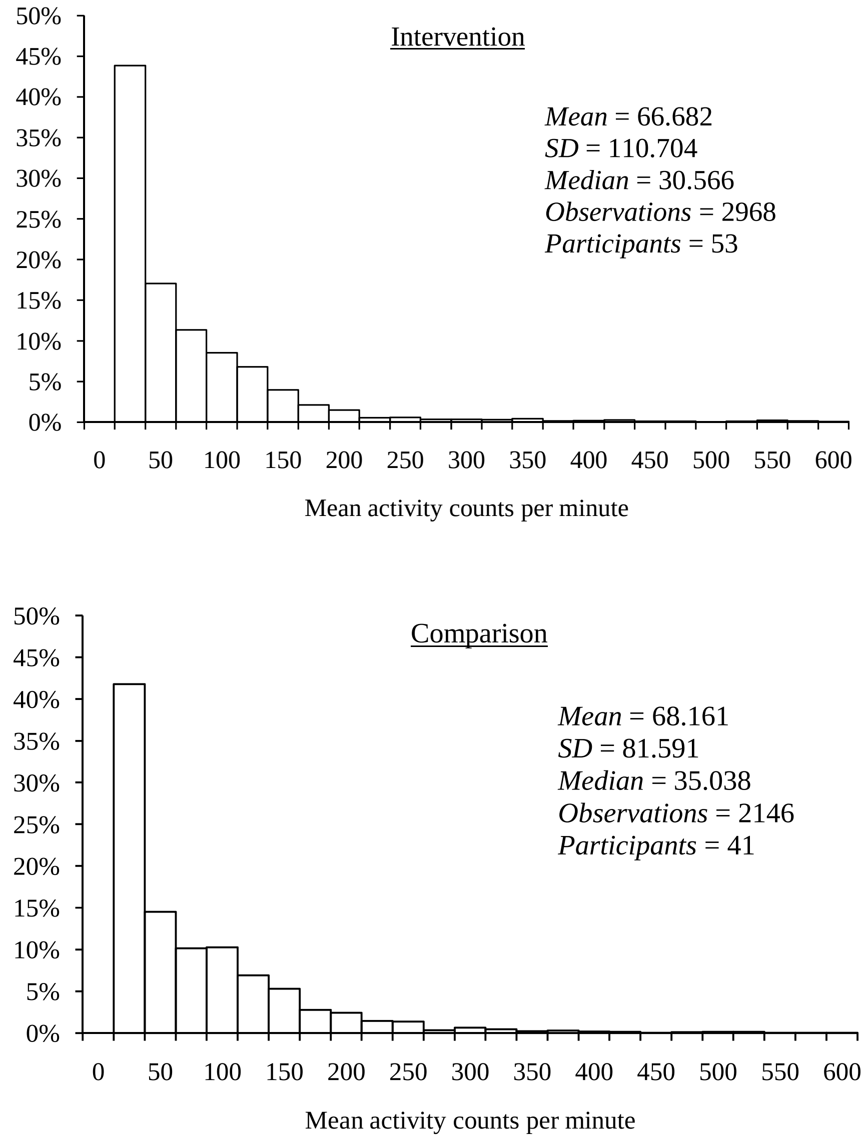


Figure VI-15. Histograms of mean activity counts for all sleep observations by treatment condition.

3. Profile of Mood States

The study examined how Profile of Mood States (POMS) factor scores related to the treatment condition over the course of Basic Combat Training while accounting for potential covariates and the known differences between the study groups. However, any approach to modeling the POMS factor scores needed to address several issues. First, a

MANCOVA of the pre-study POMS factor scores with treatment condition as a fixed effect and age, caffeine and tobacco use, component, GT score, firearm use, fitness factors (BMI and exercise frequency), NEO personality component scores, RSES score, sex, and sleep factors (ESS and PSQI scores) as covariates found a significant effect for treatment condition (Wilks' $\lambda = 0.769$, $F_{6,367} = 18.393$, $p < 0.001$). An examination of the univariate ANCOVAs showed that there were significant fixed effects for treatment condition on T-factor (tension-anxiety) scores ($F_{1,372} = 42.094$, $p < 0.001$), D-factor (depression-dejection) scores ($F_{1,372} = 30.305$, $p < 0.001$), A-factor (anger-hostility) scores ($F_{1,372} = 39.278$, $p < 0.001$), V-factor (vigor-activity) scores ($F_{1,372} = 6.961$, $p = 0.009$), F-factor (fatigue-inertia) scores ($F_{1,372} = 100.803$, $p < 0.001$), and C-factor (confusion-bewilderment) scores ($F_{1,372} = 22.397$, $p < 0.001$). It was clearly observed from Figure VI-16 that the pre-study POMS factor scores, prior to any exposure to the treatment, differed between the study groups.

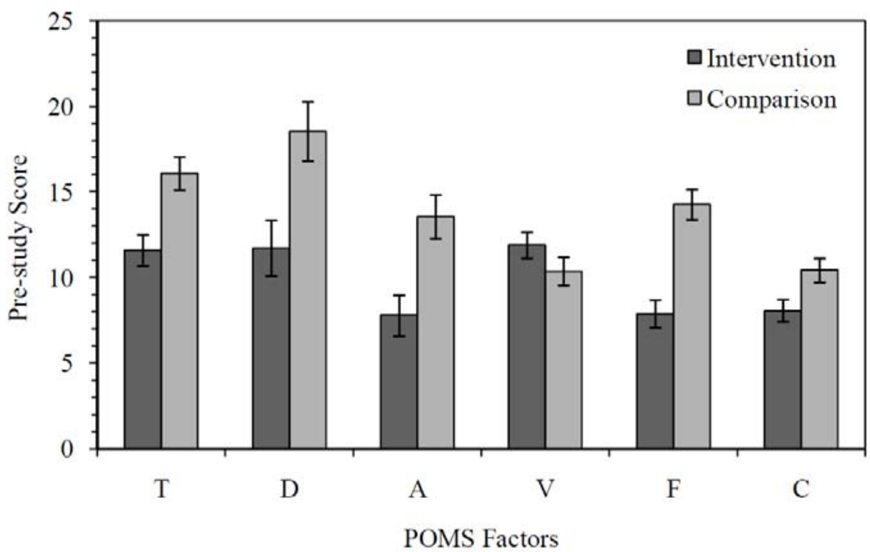


Figure VI-16. Comparison of estimated marginal means and associated 95% confidence intervals for pre-study POMS factor scores by study group.

These results suggested that the two study groups were not directly comparable at baseline in terms of subjective mood. This issue was remedied by calculating the "delta

from baseline" score for each factor—that is, subtracting a participant's pre-study POMS factor score from all their subsequent POMS factor scores. This subtraction had the effect of making all participants' pre-study POMS factor scores zero, while still preserving the magnitude and directionality of variations in their subsequent POMS factor scores. Another issue was the observation that most participants (70.4%) did not have a POMS questionnaire for every week of training. This issue was addressed by analyzing the POMS dataset as a repeated cross-section design rather than a within-participant repeated measures design and using a 1% significance level to counter the resulting increased power of statistical tests.

A MANCOVA of the POMS factor delta from baseline scores was accomplished using treatment condition, week, and chronotype as fixed effects and age, caffeine and tobacco use, component, firearm use, fitness factors (BMI and exercise frequency), GT score, NEO personality component scores, RSES score, sex, and sleep factors (ESS and PSQI scores) as covariates. Table VI-11 summarizes the results of the multivariate tests. There were significant fixed effects for treatment condition, week, and chronotype as well as significant interaction effects between treatment condition and both week and chronotype. With the exception of exercise frequency, firearm use, NEO extraversion component score, and RSES score, there were significant effects for all the measured covariates.

Table VI-11. Multivariate tests for POMS delta from baseline scores.

Source	Wilks' λ	*F*	df1	df2	*p*	η^2
Condition	0.992	4.261	6	3037	<0.001*	0.008
Week	0.944	3.694	48	14947	<0.001*	0.010
Chronotype	0.984	4.217	12	6074	<0.001*	0.008
Condition x Week	0.974	1.673	48	14947	0.002*	0.004
Condition x Chronotype	0.990	2.628	12	6074	0.002*	0.005
Chronotype x Week	0.985	0.466	96	17213	1.000	0.002
Condition x Chronotype x Week	0.981	0.617	96	17213	0.999	0.003
Age	0.967	17.008	6	3037	<0.001*	0.033
Body mass index	0.980	10.084	6	3037	<0.001*	0.020
Caffeine use (referent no)	0.981	9.842	6	3037	<0.001*	0.019
Component (referent regular)	0.989	5.812	6	3037	<0.001*	0.011
Epworth Sleepiness Scale	0.956	23.510	6	3037	<0.001*	0.044
Exercise frequency	0.995	2.628	6	3037	0.015	0.005
Firearm use (referent no)	0.996	1.951	6	3037	0.069	0.004
GT score	0.968	16.607	6	3037	<0.001*	0.032
NEO-FFI						
Neuroticism	0.966	17.934	6	3037	<0.001*	0.034
Extraversion	0.995	2.318	6	3037	0.031	0.005
Openness to experience	0.985	7.631	6	3037	<0.001*	0.015
Agreeableness	0.973	14.192	6	3037	<0.001*	0.027
Conscientiousness	0.982	9.075	6	3037	<0.001*	0.018
Pittsburgh Sleep Quality Index	0.984	8.108	6	3037	<0.001*	0.016
RSES	0.995	2.583	6	3037	0.017	0.005
Sex (referent male)	0.973	13.883	6	3037	<0.001*	0.027
Tobacco use (referent no)	0.988	6.158	6	3037	<0.001*	0.012

*Significant at ≤ 0.01 level.

Notes: GT score = General technical aptitude score; NEO-FFI = NEO Five-Factor Inventory; RSES = Response to Stressful Experiences Scale.

a. Tension-Anxiety (T) Factor

Table VI-12 provides the results of the relevant univariate tests of between-participant effects for the POMS T-factor delta from baseline scores. There was no significant fixed effect for treatment condition or chronotype.

Table VI-12. Univariate tests of between-participant effects for POMS T-factor delta from baseline scores.

Source	MS	df	*F*	*p*	η^2
Condition	60.636	1	1.359	0.244	<0.001
Week	335.619	8	7.521	<0.001*	0.019
Chronotype	31.538	2	0.707	0.493	<0.001
Condition x Week	78.945	8	1.769	0.078	0.005
Condition x Chronotype	49.363	2	1.106	0.331	0.001
Age	555.040	1	12.439	<0.001*	0.004
Body mass index	1243.017	1	27.857	<0.001*	0.009
Caffeine use (referent no)	814.800	1	18.260	<0.001*	0.006
Component (referent regular)	219.848	1	4.927	0.027	0.002
Epworth Sleepiness Scale	124.464	1	2.789	0.095	0.001
GT score	1474.994	1	33.055	<0.001*	0.011
NEO-FFI					
Neuroticism	1379.661	1	30.919	<0.001*	0.010
Openness to experience	80.314	1	1.800	0.180	0.001
Agreeableness	14.529	1	0.326	0.568	<0.001
Conscientiousness	20.671	1	0.463	0.496	<0.001
Pittsburgh Sleep Quality Index	762.339	1	17.084	<0.001*	0.006
Sex (referent male)	298.227	1	6.683	0.010*	0.002
Tobacco use (referent no)	706.302	1	15.829	<0.001*	0.005
Error	44.622	3042			

*Significant at ≤ 0.01 level.
Notes: GT score = General technical aptitude score; MS = Mean square; NEO-FFI = NEO Five-Factor Inventory.

There was a significant fixed effect for week (Figure VI-17), with the main pairwise differences occurring between week 1 versus weeks 4–7 and week 9 ($p \leq 0.001$) and between week 3 versus week 6 ($p = 0.006$).

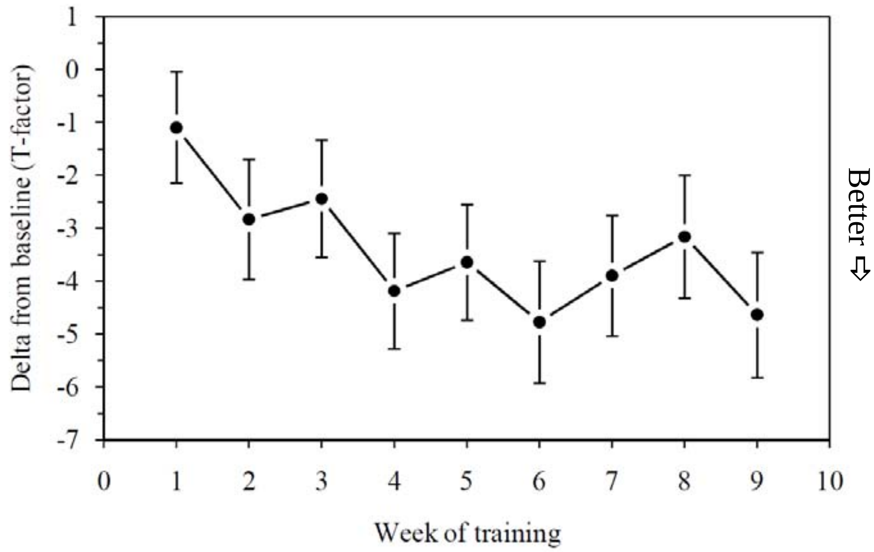


Figure VI-17. Estimated marginal means for POMS T-factor delta from baseline scores by week of training (error bars are for 99% confidence intervals).

There was no significant interaction effect between treatment condition and either week or chronotype. Thus, the general trend was for T-factor scores to decrease during the first six weeks of training followed by a spike in T-factor scores during weeks 7–8. Significant covariates included age, BMI, caffeine and tobacco use, GT score, NEO neuroticism component score, PSQI score, and sex, but only GT score, and neuroticism had effect sizes of at least 1% as measured using eta squared.

b. Depression-Dejection (D) Factor

Table VI-13 provides the results of the univariate tests of between-participant effects for the POMS D-factor delta from baseline scores. Again, there was no significant fixed effect for treatment condition or chronotype.

Table VI-13. Univariate tests of between-participant effects for POMS D-factor delta from baseline scores.

Source	MS	df	*F*	*p*	η^2
Condition	132.618	1	0.989	0.320	<0.001
Week	1208.472	8	9.015	<0.001*	0.023
Chronotype	299.645	2	2.235	0.107	0.001
Condition x Week	158.458	8	1.182	0.306	0.003
Condition x Chronotype	245.889	2	1.834	0.160	0.001
Age	1014.065	1	7.565	0.006*	0.002
Body mass index	5334.391	1	39.793	<0.001*	0.013
Caffeine use (referent no)	2135.415	1	15.930	<0.001*	0.005
Component (referent regular)	146.044	1	1.089	0.297	<0.001
Epworth Sleepiness Scale	0.044	1	0.000	0.985	<0.001
GT score	856.795	1	6.391	0.012	0.002
NEO-FFI					
Neuroticism	6150.683	1	45.882	<0.001*	0.015
Openness to experience	577.989	1	4.312	0.038	0.001
Agreeableness	2046.344	1	15.265	<0.001*	0.005
Conscientiousness	708.772	1	5.287	0.022	0.002
Pittsburgh Sleep Quality Index	233.218	1	1.740	0.187	0.001
Sex (referent male)	165.777	1	1.237	0.266	<0.001
Tobacco use (referent no)	518.436	1	3.867	0.049	0.001
Error	134.054	3042			

*Significant at ≤ 0.01 level.
Notes: GT score = General technical aptitude score; MS = Mean square; NEO-FFI = NEO Five-Factor Inventory.

There was a significant fixed effect for week (Figure VI-18), with pairwise differences occurring between week 1 versus weeks 4–9 ($p \leq 0.002$), week 2 versus week 9 ($p = 0.001$), and week 3 versus week 9 ($p = 0.003$).

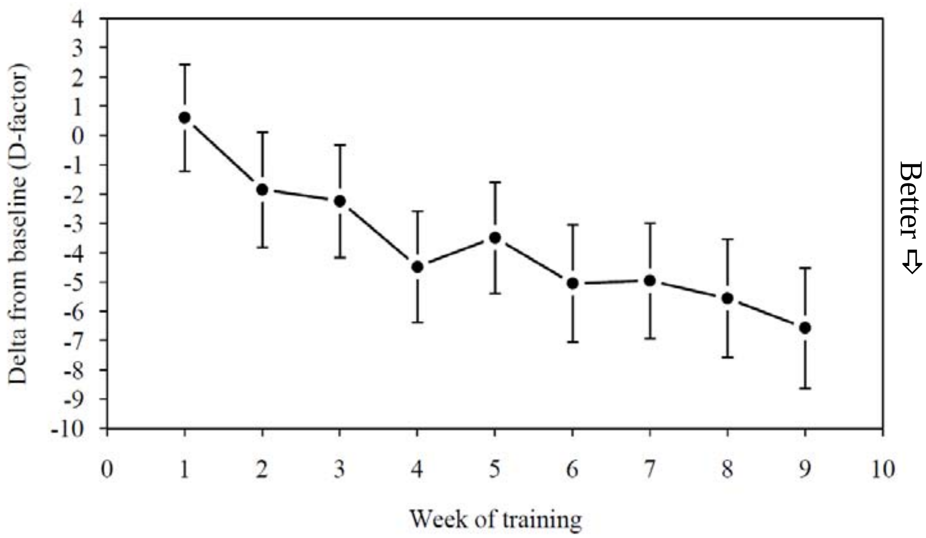


Figure VI-18. Estimated marginal means for POMS D-factor delta from baseline scores by week of training (error bars are for 99% confidence intervals).

There was no significant interaction effect between treatment condition and either week or chronotype. Thus, the general trend was for D-factor scores to decrease during the course of training, with lower scores meaning less of a depressed mood. Significant covariates included age, BMI, caffeine use, and NEO neuroticism and agreeableness component scores, but only BMI and neuroticism had effect sizes of at least 1%.

c. Anger-Hostility (A) Factor

Table VI-14 provides the results of the univariate tests of between-participant effects for the POMS A-factor delta from baseline scores. There was no significant fixed effect for treatment condition or chronotype.

Table VI-14. Univariate tests of between-participant effects for POMS A-factor delta from baseline scores.

Source	MS	df	F	p	η^2
Condition	5.447	1	0.062	0.803	<0.001
Week	718.227	8	8.172	<0.001*	0.021
Chronotype	118.510	2	1.348	0.260	0.001
Condition x Week	235.186	8	2.676	0.006*	0.007
Condition x Chronotype	200.591	2	2.282	0.102	0.001
Age	1553.745	1	17.679	<0.001*	0.006
Body mass index	1822.769	1	20.740	<0.001*	0.007
Caffeine use (referent no)	538.882	1	6.131	0.013	0.002
Component (referent regular)	38.695	1	0.440	0.507	<0.001
Epworth Sleepiness Scale	34.238	1	0.390	0.533	<0.001
GT score	1301.170	1	14.805	<0.001*	0.005
NEO-FFI					
Neuroticism	176.461	1	2.008	0.157	0.001
Openness to experience	1270.906	1	14.461	<0.001*	0.005
Agreeableness	7.572	1	0.086	0.769	<0.001
Conscientiousness	252.873	1	2.877	0.090	0.001
Pittsburgh Sleep Quality Index	158.508	1	1.804	0.179	0.001
Sex (referent male)	3035.072	1	34.533	<0.001*	0.011
Tobacco use (referent no)	963.306	1	10.961	0.001*	0.004
Error	87.888	3042			

*Significant at ≤ 0.01 level.

Notes: GT score = General technical aptitude score; MS = Mean square; NEO-FFI = NEO Five-Factor Inventory.

There was a significant fixed effect for week (Figure VI-19), with the pairwise differences occurring between week 1 versus week 4 and weeks 6–9 ($p \leq 0.005$), week 2 versus week 9 ($p = 0.001$), week 3 versus week 9 ($p < 0.001$), and week 5 versus week 9 ($p = 0.002$).

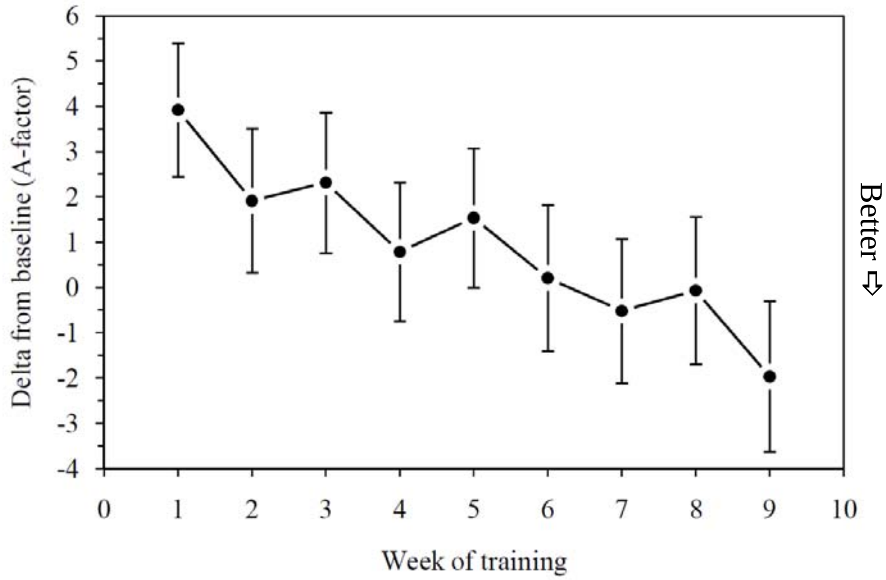


Figure VI-19. Estimated marginal means for POMS A-factor delta from baseline scores by week of training (error bars are for 99% confidence intervals).

There was a significant interaction effect between treatment condition and week (Figure VI-20), but not between treatment condition and chronotype.

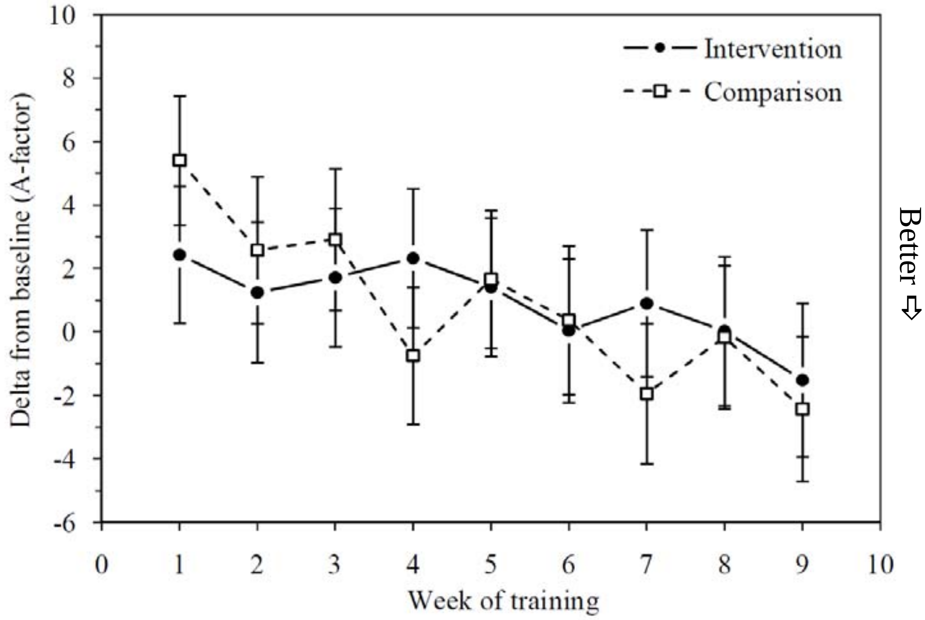


Figure VI-20. Estimated marginal means for POMS A-factor delta from baseline scores by treatment condition and week of training (error bars are for 99% confidence intervals).

Thus, the comparison group started out with higher A-factor delta from baseline scores but had a greater rate of decrease in scores over training as compared to the intervention group. Significant covariates included age, BMI, GT score, NEO openness to experience component score, sex, and tobacco use, but only sex had an effect size of at least 1%.

d. Vigor-Activity (V) Factor

Table VI-15 provides the results of the univariate tests of between-participant effects for the POMS V-factor delta from baseline scores. There was a significant fixed effect for treatment condition with an estimated marginal mean score for the intervention group of 1.229 (99% CI: 0.830, 1.628) versus 0.098 (99% CI: –0.347, 0.543) for the comparison group. That is, controlling for other variables, the intervention group exhibited a mood of greater vigorousness and ebullience and higher energy than the comparison group.

Table VI-15. Univariate tests of between-participant effects for POMS V-factor delta from baseline scores.

Source	MS	df	*F*	*p*	η^2
Condition	394.489	1	10.232	0.001*	0.003
Week	17.975	8	0.466	0.881	0.001
Chronotype	574.906	2	14.911	<0.001*	0.010
Condition x Week	78.426	8	2.034	0.039	0.005
Condition x Chronotype	94.740	2	2.457	0.086	0.002
Age	3039.636	1	78.838	<0.001*	0.025
Body mass index	571.114	1	14.813	<0.001*	0.005
Caffeine use (referent no)	377.387	1	9.788	0.002*	0.003
Component (referent regular)	494.366	1	12.822	<0.001*	0.004
Epworth Sleepiness Scale	2844.343	1	73.773	<0.001*	0.024
GT score	1283.601	1	33.292	<0.001*	0.011
NEO-FFI					
Neuroticism	1037.429	1	26.907	<0.001*	0.009
Openness to experience	479.607	1	12.439	<0.001*	0.004
Agreeableness	224.950	1	5.834	0.016	0.002
Conscientiousness	378.944	1	9.829	0.002*	0.003
Pittsburgh Sleep Quality Index	395.210	1	10.250	0.001*	0.003
Sex (referent male)	561.431	1	14.562	<0.001*	0.005
Tobacco use (referent no)	40.373	1	1.047	0.306	<0.001
Error	38.555	3042			

*Significant at ≤ 0.01 level.

Notes: GT score = General technical aptitude score; MS = Mean square; NEO-FFI = NEO Five-Factor Inventory.

There was no significant fixed effect for week, but there was a significant effect for chronotype (Figure IV-21), with the main pairwise difference occurring between evening and indeterminate chronotypes ($p < 0.001$).

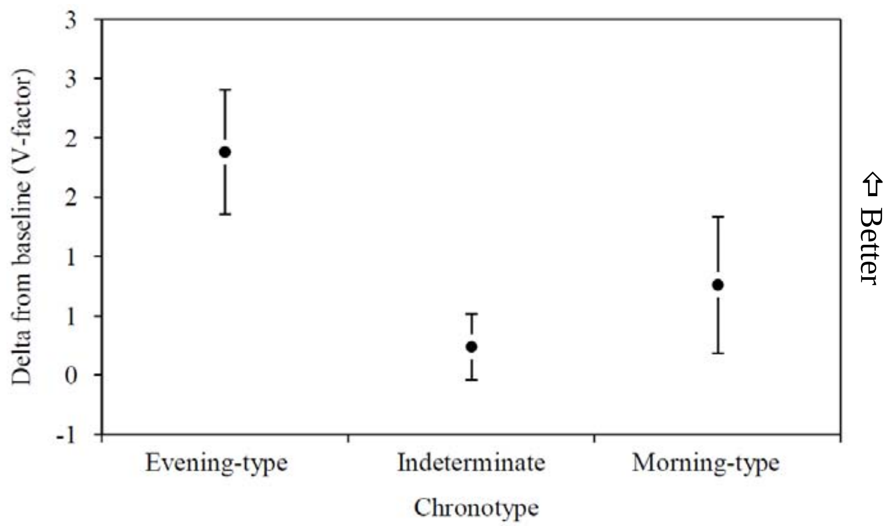


Figure VI-21. Estimated marginal means for POMS V-factor delta from baseline scores by chronotype (error bars are for 99% confidence intervals).

Significant covariates included age, BMI, caffeine use, component, ESS score, GT score, NEO (neuroticism, openness to experience, and agreeableness component scores), PSQI score, and sex. Only age, ESS score, and GT score had effect sizes of at least 1%.

e. Fatigue-Inertia (F) Factor

Table VI-16 provides the results of the univariate tests of between-participant effects for the POMS F-factor delta from baseline scores. There were no significant fixed effects of either treatment condition or chronotype. However, there was a significant fixed effect of week as well as a significant interaction effect between treatment condition and week.

Table VI-16. Univariate tests of between-participant effects for POMS F-factor delta from baseline scores.

Source	MS	df	*F*	*p*	η^2
Condition	119.754	1	3.092	0.079	0.001
Week	401.350	8	10.362	<0.001*	0.027
Chronotype	23.846	2	0.616	0.540	<0.001
Condition x Week	163.341	8	4.217	<0.001*	0.011
Condition x Chronotype	111.529	2	2.880	0.056	0.002
Age	1100.898	1	28.424	<0.001*	0.009
Body mass index	1451.967	1	37.488	<0.001*	0.012
Caffeine use (referent no)	112.819	1	2.913	0.088	0.001
Component (referent regular)	16.907	1	0.437	0.509	<0.001
Epworth Sleepiness Scale	2118.381	1	54.694	<0.001*	0.018
GT score	753.970	1	19.467	<0.001*	0.006
NEO-FFI					
Neuroticism	627.055	1	16.190	<0.001*	0.005
Openness to experience	8.629	1	0.223	0.637	<0.001
Agreeableness	1108.981	1	28.633	<0.001*	0.009
Conscientiousness	899.462	1	23.223	<0.001*	0.008
Pittsburgh Sleep Quality Index	33.364	1	0.861	0.353	<0.001
Sex (referent male)	472.120	1	12.190	<0.001*	0.004
Tobacco use (referent no)	33.269	1	0.859	0.354	<0.001
Error	38.731	3042			

*Significant at ≤ 0.01 level.

Notes: GT score = General technical aptitude score; MS = Mean square; NEO-FFI = NEO Five-Factor Inventory.

For the fixed effect of week (Figure VI-22), the pairwise differences occurred between week 1 versus week 4 and weeks 6–9 ($p < 0.001$); week 2 versus week 7 ($p = 0.009$); week 3 versus weeks 6, 7, and 9 ($p \leq 0.009$); and week 5 versus weeks 4, 6–7, and 9 ($p \leq 0.005$).

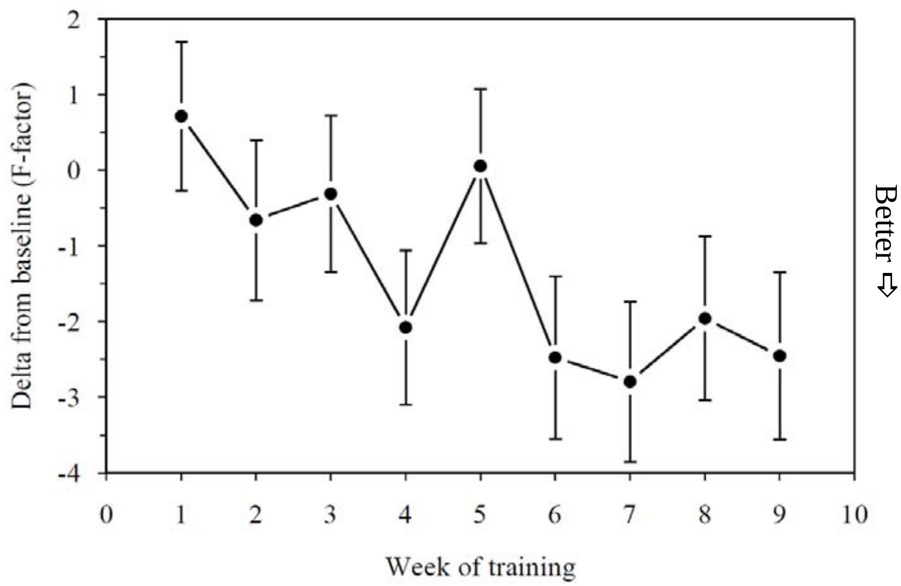


Figure VI-22. Estimated marginal means for POMS F-factor delta from baseline scores by week of training (error bars are for 99% confidence intervals).

In terms of the significant interaction effect (Figure VI-23), the comparison group started out with a higher mean F-factor score but had a greater rate of decrease in scores over training as compared to the intervention group.

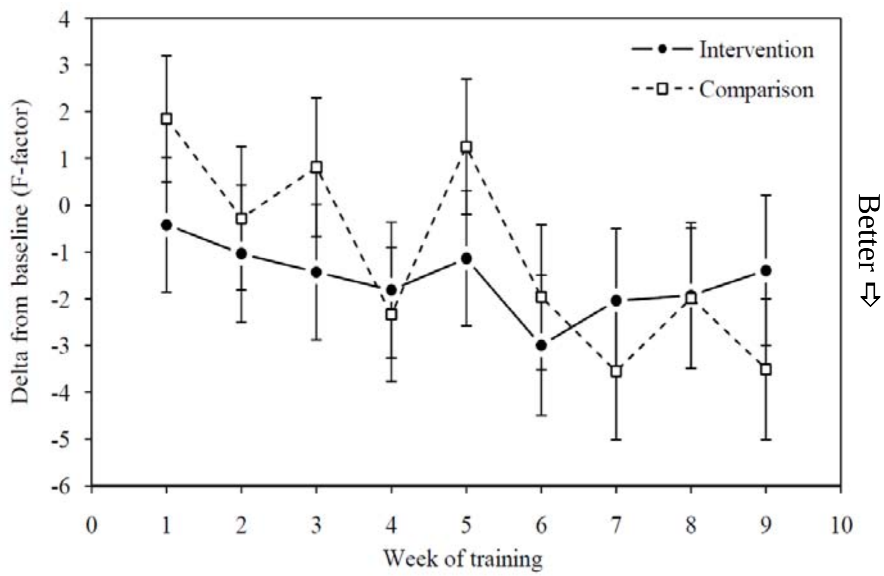


Figure VI-23. Estimated marginal means for POMS F-factor delta from baseline scores by treatment condition and week of training (error bars are for 99% confidence intervals).

Significant covariates included age, BMI, ESS score, GT score, NEO (neuroticism, agreeableness, and conscientiousness component scores), and sex. Only BMI and ESS score had effect sizes of at least 1%.

f. Confusion-Bewilderment (C) Factor

Table VI-17 provides the results of the univariate tests of between-participant effects for the POMS C-factor delta from baseline scores. There was no significant fixed effect for treatment condition or chronotype.

Table VI-17. Univariate tests of between-participant effects for POMS C-factor delta from baseline scores.

Source	MS	df	F	p	η^2
Condition	26.964	1	1.117	0.291	<0.001
Week	274.662	8	11.383	<0.001*	0.029
Chronotype	5.940	2	0.246	0.782	<0.001
Condition x Week	27.565	8	1.142	0.331	0.003
Condition x Chronotype	27.612	2	1.144	0.319	0.001
Age	30.062	1	1.246	0.264	<0.001
Body mass index	790.474	1	32.760	<0.001*	0.011
Caffeine use (referent no)	38.958	1	1.615	0.204	0.001
Component (referent regular)	274.152	1	11.362	0.001*	0.004
Epworth Sleepiness Scale	248.181	1	10.286	0.001*	0.003
GT score	72.149	1	2.990	0.084	0.001
NEO-FFI					
Neuroticism	181.822	1	7.535	0.006*	0.002
Openness to experience	2.737	1	0.113	0.736	<0.001
Agreeableness	92.860	1	3.848	0.050	0.001
Conscientiousness	286.123	1	11.858	0.001*	0.004
Pittsburgh Sleep Quality Index	449.225	1	18.618	<0.001*	0.006
Sex (referent male)	57.315	1	2.375	0.123	0.001
Tobacco use (referent no)	446.382	1	18.500	<0.001*	0.006
Error	24.129	3042			

*Significant at ≤ 0.01 level.

Notes: GT score = General technical aptitude score; MS = Mean square; NEO-FFI = NEO Five-Factor Inventory.

There was a significant fixed effect for week (Figure VI-24), with pairwise differences occurring between week 1 versus weeks 3–9 ($p < 0.006$) and week 2 versus weeks 6–9 ($p \leq 0.005$).

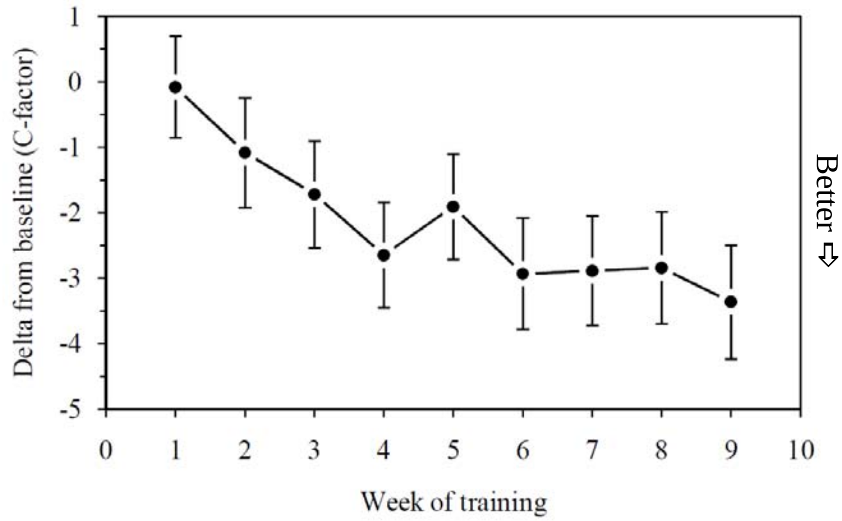


Figure VI-24. Estimated marginal means for POMS C-factor delta from baseline scores by week of training (error bars are for 99% confidence intervals).

There was no significant interaction effect between treatment condition and either week or chronotype. Thus, the trend was for C-factor scores to decrease during the course of training. Significant covariates included BMI, component, ESS score, NEO neuroticism and conscientiousness component scores, PSQI score, and tobacco use, but only BMI had an effect size of at least 1%.

g. Total Mood Disturbance Score

A total mood disturbance (TMD) score was obtained from the POMS by simply summing the scores across all six factors while negatively weighting vigor. Accordingly, the TMD score provides a single global estimate of affective state (McNair & Heuchert, 2005). An ANCOVA of TMD delta from baseline scores was accomplished using treatment condition, week, and chronotype as fixed effects and age, caffeine and tobacco use, component, firearm use, fitness factors (BMI and exercise frequency), GT

score, personality component scores, RSES score, sex, and sleep factors (ESS and PSQI scores) as covariates (Table VI-18).

Table VI-18. Univariate tests for Total Mood Disturbance delta from baseline scores.

Source	MS	df	*F*	*p*	η^2
Condition	253.538	1	0.221	0.638	<0.001
Week	12915.545	8	11.276	<0.001*	0.029
Chronotype	1400.551	2	1.223	0.295	0.001
Condition x Week	3306.386	8	2.887	0.003*	0.008
Condition x Chronotype	2040.045	2	1.781	0.169	0.001
Chronotype x Week	839.027	16	0.733	0.763	0.004
Condition x Chronotype x Week	1137.775	16	0.993	0.461	0.005
Age	36498.019	1	31.865	<0.001*	0.010
Body mass index	58619.151	1	51.178	<0.001*	0.017
Caffeine use (referent no)	5566.435	1	4.860	0.028	0.002
Component (referent regular)	153.641	1	0.134	0.714	<0.001
Epworth Sleepiness Scale	17536.589	1	15.311	<0.001*	0.005
Exercise frequency	2809.579	1	2.453	0.117	0.001
Firearm use (referent no)	557.135	1	0.486	0.486	<0.001
GT score	10973.626	1	9.581	0.002*	0.003
NEO-FFI					
Neuroticism	40202.835	1	35.100	<0.001*	0.011
Extraversion	2535.015	1	2.213	0.137	0.001
Openness to experience	5919.692	1	5.168	0.023	0.002
Agreeableness	9377.554	1	8.187	0.004*	0.003
Conscientiousness	10897.472	1	9.514	0.002*	0.003
Pittsburgh Sleep Quality Index	2656.198	1	2.319	0.128	0.001
RSES	7.987	1	0.007	0.933	<0.001
Sex (referent male)	8096.891	1	7.069	0.008*	0.002
Tobacco use (referent no)	12866.257	1	11.233	0.001*	0.004
Error	1145.388	3039			

*Significant at ≤ 0.01 level.
Notes: GT score = General technical aptitude score; MS = Mean square; NEO-FFI = NEO Five-Factor Inventory; RSES = Response to Stressful Experiences Scale.

There were no significant fixed effects for treatment condition or chronotype. However, there was a significant fixed effect for week as well as a significant interaction effect between treatment condition and week. For the fixed effect of week (Figure VI-25), pairwise differences occurred between week 1 versus weeks 4–9 ($p \leq 0.004$), week 2 versus week 9 ($p < 0.001$), week 3 versus week 9 ($p < 0.001$), and week 5 versus week 9 ($p = 0.007$).

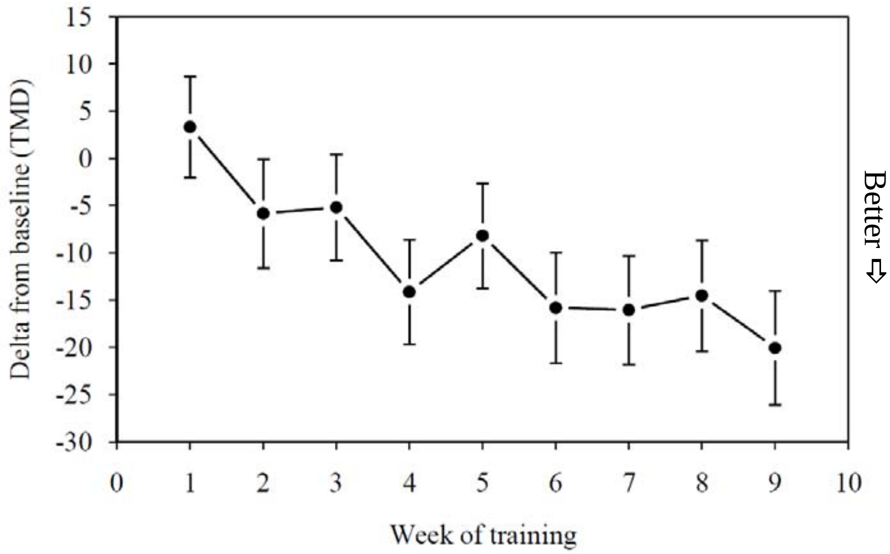


Figure VI-25. Estimated marginal means for POMS Total Mood Disturbance (TMD) delta from baseline scores by week of training (error bars are for 99% confidence intervals).

As shown in Figure VI-26, the comparison group started out with a higher mean TMD score but had a greater rate of decrease in scores over the course of training relative to the intervention group.

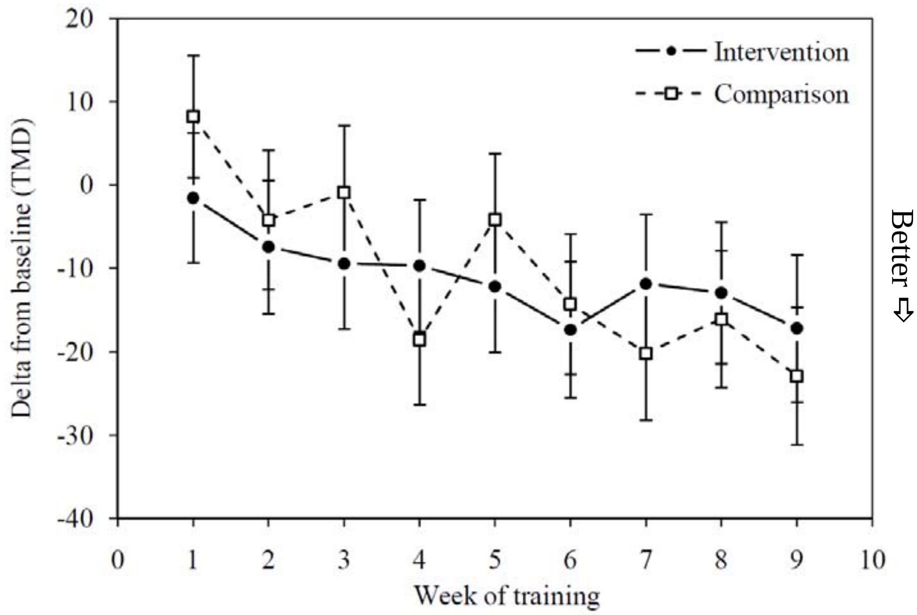


Figure VI-26. Estimated marginal means for POMS Total Mood Disturbance (TMD) delta from baseline scores by treatment condition and week of training (error bars are for 99% confidence intervals).

Significant covariates included age, BMI, ESS score, GT score, NEO (neuroticism, agreeableness, and conscientiousness component scores), sex, and tobacco use. Only age, BMI, and neuroticism had effect sizes of at least 1%.

h. Actigraphy Subsample

The analysis of the POMS data was repeated for the subsample of participants for which actigraphy data was available. The same analytic approach was used with the exception that weekly average hours slept was used as the covariate. Table VI-19 summarizes the results of the multivariate tests. There was no significant fixed effect of treatment condition or week, but there was a significant fixed effect of chronotype as well as a significant interaction effect between treatment condition and

chronotype. There was also a significant multivariate effect of the covariate, weekly average hours slept, but the covariate was not significant in any of the subsequent univariate tests.

Table VI-19. Multivariate tests for POMS delta from baseline scores for actigraphy subsample.

Source	Wilks' λ	*F*	df1	df2	*P*	η^2
Condition	0.989	1.258	6	686	0.275	0.011
Week	0.907	1.415	48	3379	0.032	0.016
Chronotype	0.863	8.749	12	1372	<0.001*	0.071
Condition x Week	0.960	0.584	48	3379	0.990	0.007
Condition x Chronotype	0.874	7.945	12	1372	<0.001*	0.065
Chronotype x Week	0.942	0.429	96	3893	1.000	0.010
Condition x Chronotype x Week	0.947	0.394	96	3893	1.000	0.009
Average weekly sleep	0.971	3.458	6	686	0.002*	0.029

*Significant at ≤ 0.01 level.
Note: MS = Mean square.

The analysis of the respective univariate tests revealed significant fixed effects of chronotype for T-factor ($F_{2,691}$ = 15.888, p < 0.001, η^2 = 0.044), D-factor ($F_{2,691}$ = 14.710, p < 0.001, η^2 = 0.041), A-factor ($F_{2,691}$ = 9.508, p < 0.001, η^2 = 0.027), V-factor ($F_{2,691}$ = 7.730, p < 0.001, η^2 = 0.022), F-factor ($F_{2,691}$ = 16.262, p < 0.001, η^2 = 0.045), and C-factor ($F_{2,691}$ = 21.489, p < 0.001, η^2 = 0.059). In the case of T-factor, D-factor, and F-factor, pairwise differences occurred between indeterminate versus both evening and morning chronotypes; the basic pattern was as shown in Figure VI-27 for T-factor.

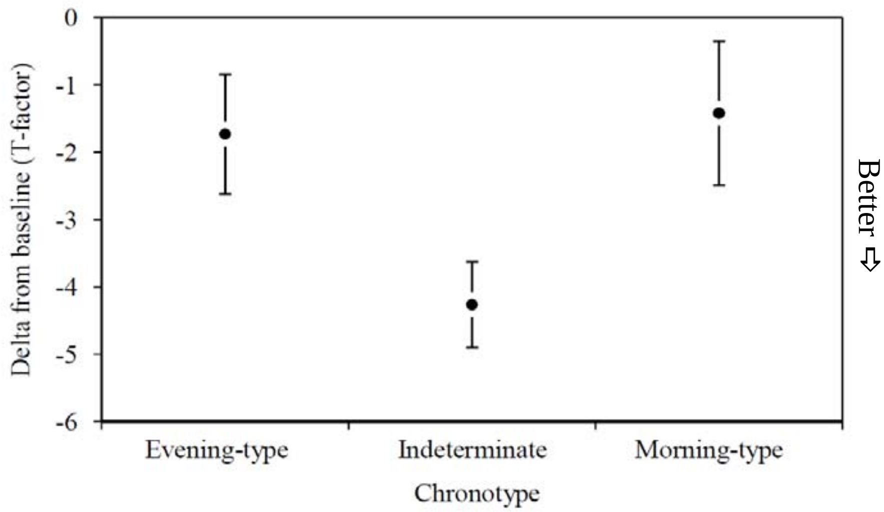


Figure VI-27. Estimated marginal means for POMS T-factor delta from baseline scores by chronotype for actigraphy subsample (error bars are for 99% confidence intervals).

For A-factor, the pairwise difference occurred between indeterminate and morning chronotypes (Figure VI-28), whereas the pairwise difference occurred between evening versus morning chronotypes for V-factor (Figure VI-29).

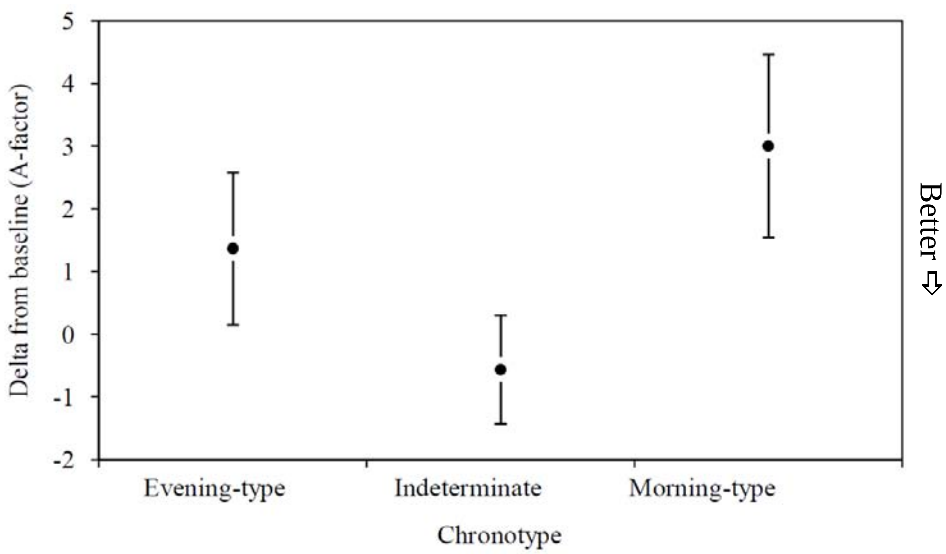


Figure VI-28. Estimated marginal means for POMS A-factor delta from baseline scores by chronotype for actigraphy subsample (error bars are for 99% confidence intervals).

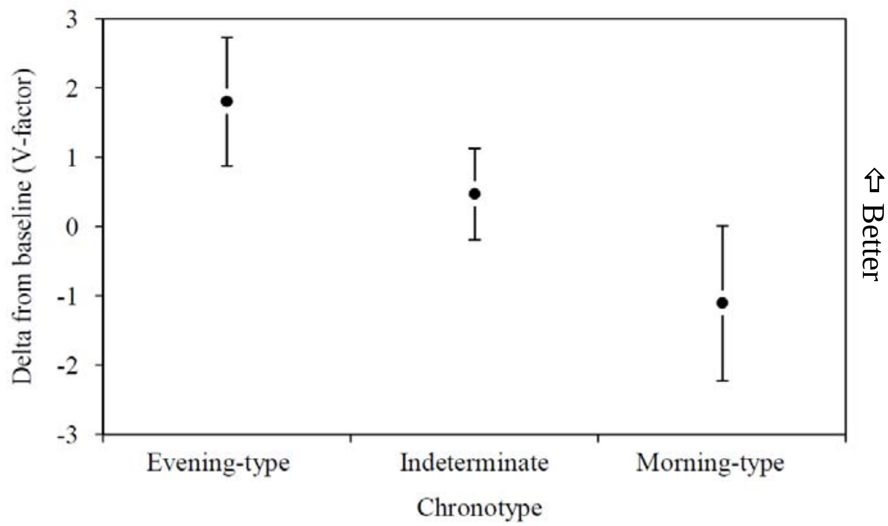


Figure VI-29. Estimated marginal means for POMS V-factor delta from baseline scores by chronotype for actigraphy subsample (error bars are for 99% confidence intervals).

In the case of C-factor, the pairwise differences occurred between evening and both indeterminate and morning chronotypes (Figure VI-30).

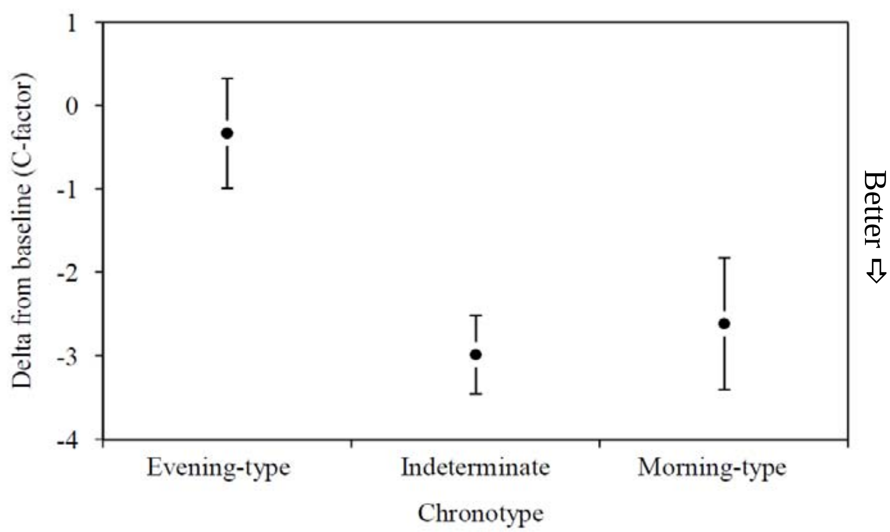


Figure VI-30. Estimated marginal means for POMS C-factor delta from baseline scores by chronotype for actigraphy subsample (error bars are for 99% confidence intervals).

The univariate tests also revealed significant interaction effects between treatment condition and chronotype for T-factor ($F_{2,691} = 14.882$, $p < 0.001$, $\eta^2 = 0.041$), D-factor ($F_{2,691} = 18.472$, $p < 0.001$, $\eta^2 = 0.051$), A-factor ($F_{2,691} = 6.264$, $p = 0.002$, $\eta^2 = 0.018$), V-factor ($F_{2,691} = 9.716$, $p < 0.001$, $\eta^2 = 0.027$), and C-factor ($F_{2,691} = 19.404$, $p < 0.001$, $\eta^2 = 0.053$). Figure VI-31 illustrates the interaction effect for D-factor; T-factor, A-factor, and C-factor followed similar patterns with evening and indeterminate chronotype participants having lower scores in the intervention group versus the comparison group, while the opposite was true for morning chronotype participants.

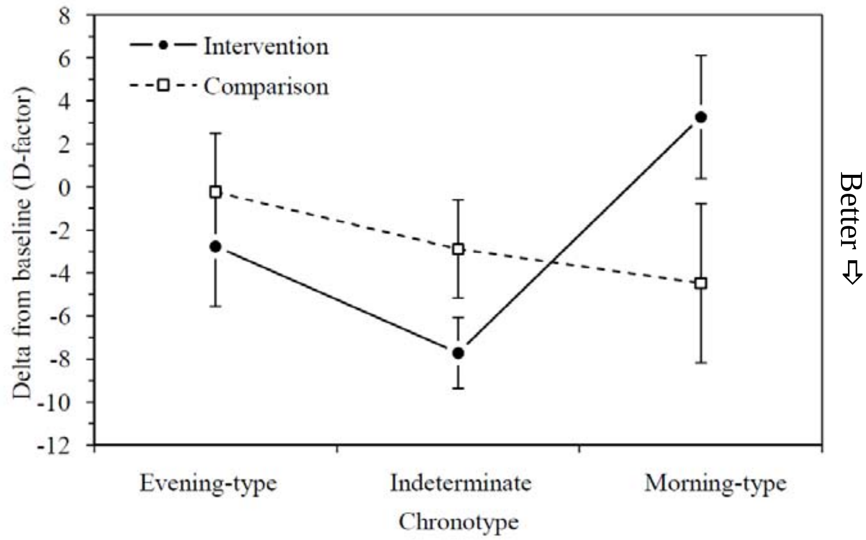


Figure VI-31. Estimated marginal means for POMS D-factor delta from baseline scores by treatment condition and chronotype for actigraphy subsample (error bars are for 99% confidence intervals).

Figure VI-32 illustrates the interaction effect for V-factor, with evening chronotype participants having lower scores in the intervention group versus the comparison group, while the opposite was true for intermediate and morning chronotype participants.

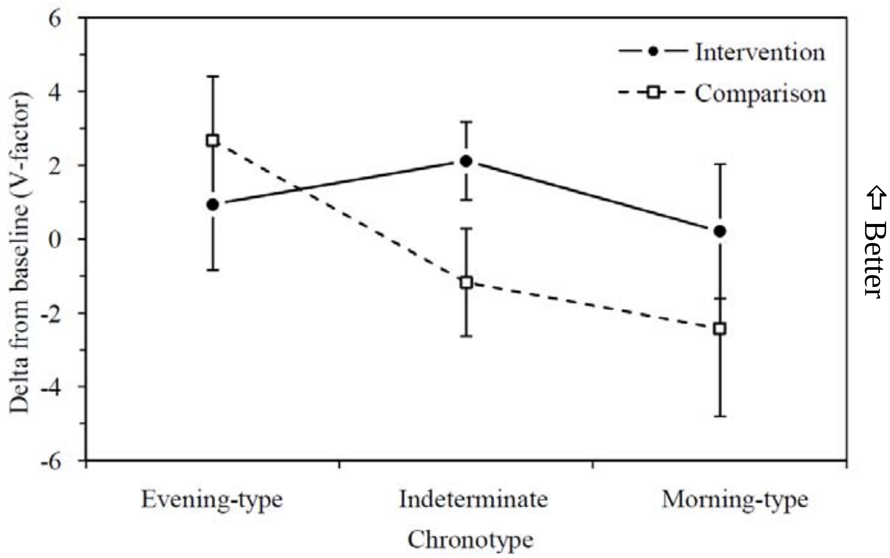


Figure VI-32. Estimated marginal means for POMS V-factor delta from baseline scores by treatment condition and chronotype for actigraphy subsample (error bars are for 99% confidence intervals).

The univariate analysis of TMD delta from baseline scores for the subsample of participants with actigraphy data (Table VI-20) showed significant fixed effects for week and chronotype but not for treatment condition.

Table VI-20. Univariate tests of between-participant effects for Total Mood Disturbance delta from baseline scores for actigraphy subsample.

Source	MS	df	*F*	*p*	η^2
Condition	2.322	1	0.003	0.960	<0.001
Week	2623.315	8	2.889	0.004*	0.032
Chronotype	16401.755	2	18.060	<0.001*	0.050
Condition x Week	1065.655	8	1.173	0.313	0.013
Condition x Chronotype	11831.703	2	13.028	<0.001*	0.036
Chronotype x Week	305.067	16	0.336	0.993	0.008
Condition x Chronotype x Week	387.332	16	0.426	0.976	0.010
Average weekly sleep	35.315	1	0.039	0.844	<0.001
Error	908.191	690			

*Significant at ≤ 0.01 level.
Note: MS = Mean square.

For the fixed effect of week (Figure VI-33), the pairwise difference occurred between week 1 versus week 9 ($p = 0.009$).

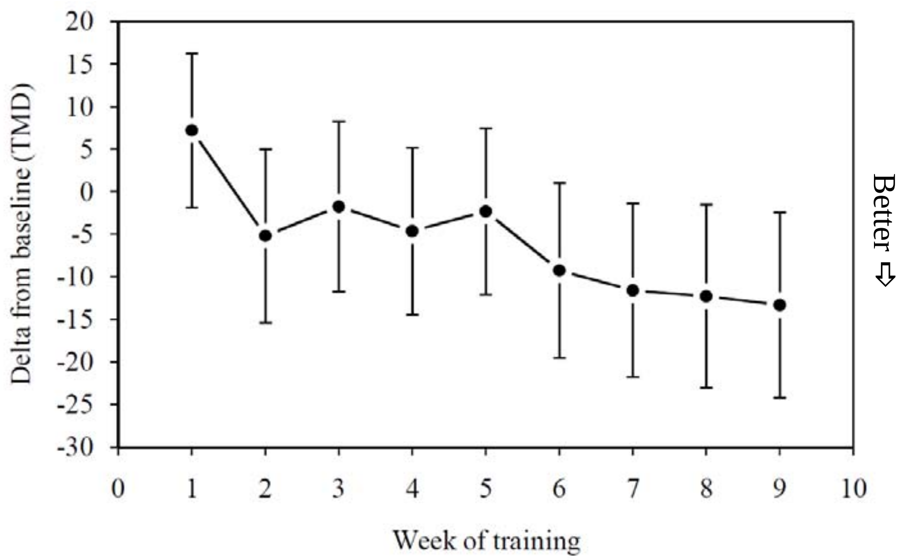


Figure VI-33. Estimated marginal means for POMS Total Mood Disturbance (TMD) delta from baseline scores by week of training for the actigraphy subsample (error bars are for 99% confidence intervals).

For the fixed effect of chronotype (Figure VI-34), pairwise differences occurred between indeterminate versus both evening and morning chronotypes ($p < 0.001$).

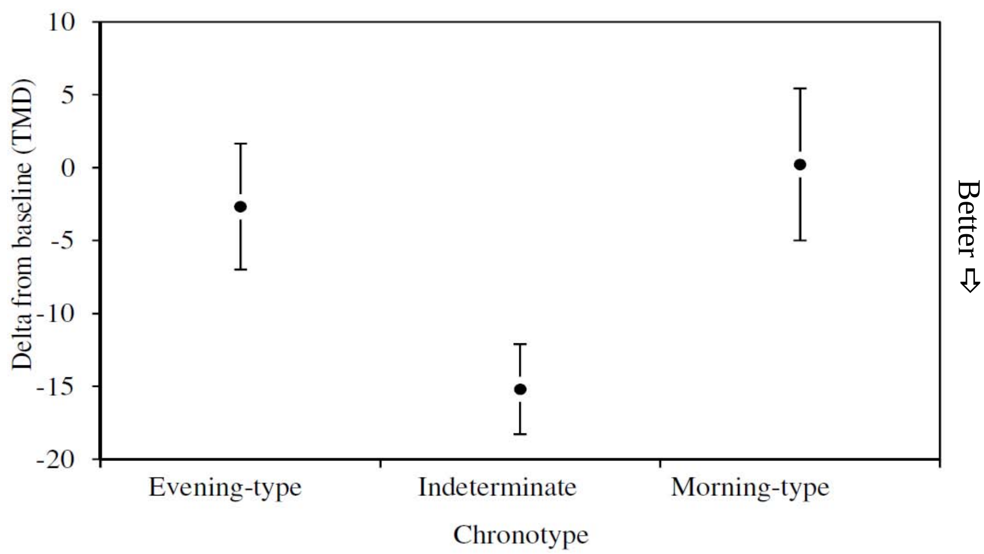


Figure VI-34. Estimated marginal means for POMS Total Mood Disturbance (TMD) delta from baseline scores by chronotype for actigraphy subsample (error bars are for 99% confidence intervals).

There was also a significant interaction effect between treatment condition and chronotype (Figure VI-35), with evening and indeterminate chronotype participants having lower scores in the intervention group versus the comparison group, while the opposite was true for morning chronotype participants. There was no significant effect of the covariate, weekly average hours slept.

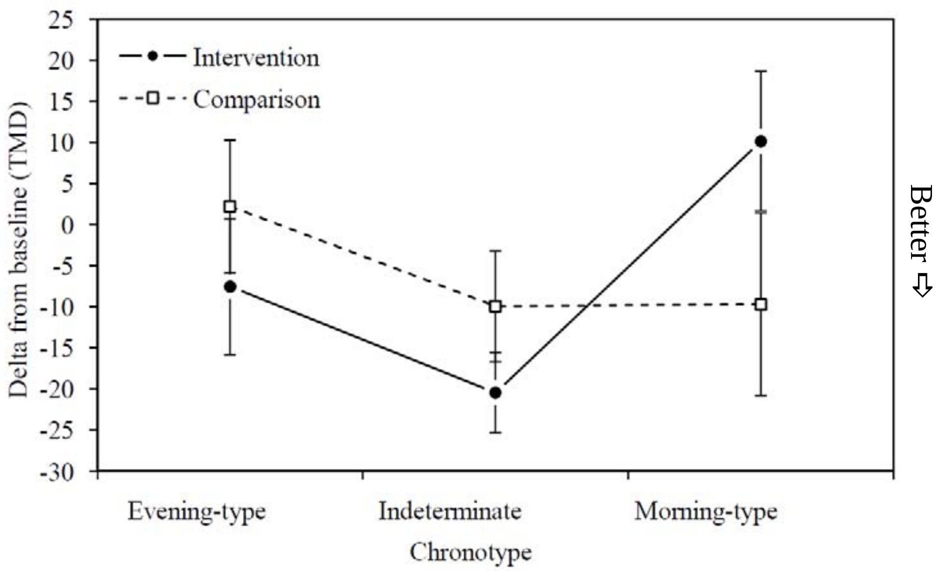


Figure VI-35. Estimated marginal means for POMS Total Mood Disturbance (TMD) delta from baseline scores by study condition and chronotype for actigraphy subsample (error bars are for 99% confidence intervals).

4. Basic Rifle Marksmanship

We assessed how participants' basic rifle marksmanship performance (on record fires) was related to treatment condition and chronotype while accounting for potential covariates. However, when the marksmanship database was received from each company, several issues needed to be addressed prior to choosing an analytical approach. First, although both companies were issued the same number of rounds per participant for basic rifle marksmanship training, each company fired those rounds at a different rate. The intervention group accomplished record fires on four separate days, while the comparison group did so on three separate days. Accordingly, there were a maximum of

four scores for each participant in the database for the intervention group and three scores per participant in the database for the comparison group. Additionally, not every participant accomplished the available maximum number of record fires.

These issues were addressed by analyzing the marksmanship scores using a simple pre/post repeated measures design in which the first recorded marksmanship score for each participant was denoted as the pre score and the last score was denoted as the post score. A repeated measures ANCOVA of marksmanship score was accomplished using practice as a within-participant effect; study condition and chronotype as fixed between-participant effects; and age, caffeine and tobacco use, component, firearm use, GT score, personality component scores, RSES score, sex, and sleep factors (ESS and PSQI scores) as covariates. In addition, given that marksmanship fundamentals were taught during the week prior to the record fires, POMS measurements from the week prior to ($t^* - 1$) and the week of (t^*) the record fires were also included as covariates.

A total of 372 participants, 201 in the intervention group (90% of the initial cohort) and 171 in the comparison group (87% of the initial cohort), had at least two observations recorded in the marksmanship databases. Tables VI-21 and VI-22 display the results for the within-participant model. Based on a 5% significance level, there was no significant within-participant effect of practice.

Table VI-21. Within-participant effects for marksmanship score.

Source	MS	df	*F*	*p*	η^2
Practice	53.799	1	2.662	0.104	0.008
Practice x Condition	196.757	1	9.737	0.002*	0.030
Practice x Chronotype	5.314	2	0.263	0.769	0.002
Practice x Condition x Chronotype	1.235	2	0.061	0.941	<0.001
Practice x Age	0.777	1	0.038	0.845	<0.001
Practice x Body mass index	14.825	1	0.734	0.392	0.002
Practice x Caffeine use (referent no)	25.043	1	1.239	0.266	0.004
Practice x Component (referent regular)	2.255	1	0.112	0.739	<0.001
Practice x Epworth Sleepiness Scale	11.565	1	0.572	0.450	0.002
Practice x Firearm use (referent no)	2.682	1	0.133	0.716	<0.001
Practice x GT score	45.644	1	2.259	0.134	0.007
Practice x NEO neuroticism	50.031	1	2.476	0.117	0.008
Practice x NEO extraversion	74.857	1	3.705	0.055	0.012
Practice x NEO openness to experience	8.837	1	0.437	0.509	0.001
Practice x NEO agreeableness	7.876	1	0.390	0.533	0.001
Practice x NEO conscientiousness	0.163	1	0.008	0.928	<0.001
Practice x PSQI	6.056	1	0.300	0.584	0.001
Practice x POMS week t^* – 1 T-factor	3.562	1	0.176	0.675	0.001
Practice x POMS week t^* – 1 D-factor	0.810	1	0.040	0.841	<0.001
Practice x POMS week t^* – 1 A-factor	27.994	1	1.385	0.240	0.004
Practice x POMS week t^* – 1 V-factor	0.865	1	0.043	0.836	<0.001
Practice x POMS week t^* – 1 F-factor	18.454	1	0.913	0.340	0.003
Practice x POMS week t^* – 1 C-factor	20.848	1	1.032	0.311	0.003

*Significant at ≤ 0.05 level.

Notes: GT score = General technical aptitude score; MS = Mean square; PSQI = Pittsburgh Sleep Quality Index; POMS = Profile of Mood States.

Table VI-22. Within-participant effects for marksmanship score (continued).

Source	MS	df	*F*	*p*	η^2
Practice x POMS week *t** T-factor	6.477	1	0.321	0.572	0.001
Practice x POMS week *t** D-factor	0.014	1	0.001	0.979	<0.001
Practice x POMS week *t** A-factor	16.824	1	0.833	0.362	0.003
Practice x POMS week *t** V-factor	8.999	1	0.445	0.505	0.001
Practice x POMS week *t** F-factor	17.276	1	0.855	0.356	0.003
Practice x POMS week *t** C-factor	83.390	1	4.127	0.043*	0.013
Practice x RSES	0.680	1	0.034	0.855	<0.001
Practice x Sex (referent male)	10.100	1	0.500	0.480	0.002
Practice x Tobacco use (referent no)	0.740	1	0.037	0.848	<0.001
Error	20.206	313			

*Significant at $\leq$ 0.05 level.
Notes: MS = Mean square; POMS = Profile of Mood States; RSES = Response to Stressful Experiences Scale.

There was a significant interaction effect between practice and treatment condition, but there was no interaction effect between practice and chronotype. Participants in the intervention group had significantly lower initial scores than participants in the comparison group, but participants in the intervention group had greater improvement in scores with practice such that their final scores were equivalent to those of participants in the comparison group (Figure VI-36). There was also a significant within-participant interaction between practice and *t** week POMS C-factor score.

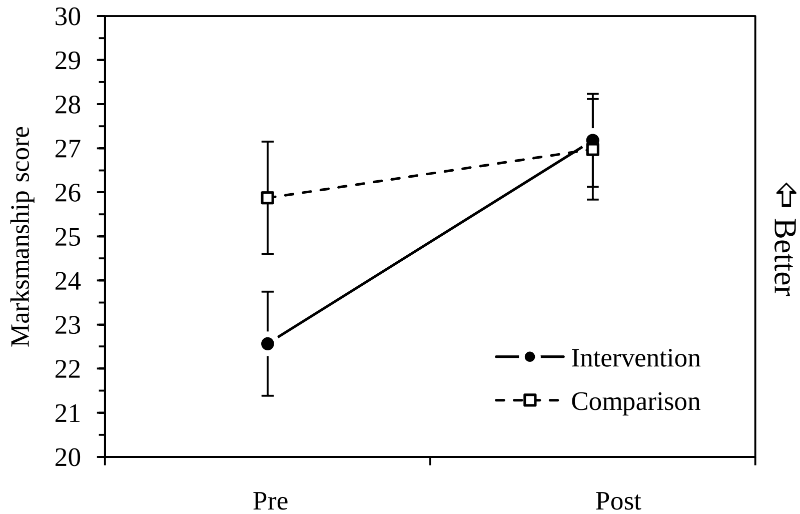


Figure VI-36. Estimated marginal means for first and last marksmanship scores by treatment condition (error bars are for 95% confidence intervals).

In terms of the between-participant model (Tables VI-23 and VI-24), there was a significant fixed effect for treatment condition, with an estimated marginal mean score for the intervention group of 24.872 (95% CI: 23.973, 25.453) versus 26.425 (95% CI: 25.772, 27.397) for the comparison group. Fixed effect of chronotype was not significant, nor was there an interaction effect between treatment condition and chronotype. The only significant covariates were prior use of firearms and sex.

Table VI-23. Between-participant effects for marksmanship score.

Source	MS	df	*F*	*p*	η^2
Condition	153.391	1	4.183	0.042*	0.013
Chronotype	5.402	2	0.147	0.863	0.001
Condition x Chronotype	43.510	2	1.186	0.307	0.008
Age	0.078	1	0.002	0.963	0.000
Body mass index	30.719	1	0.838	0.361	0.003
Caffeine use (referent no)	55.449	1	1.512	0.220	0.005
Component (referent regular)	23.717	1	0.647	0.422	0.002
Epworth Sleepiness Scale	74.759	1	2.039	0.154	0.006
Firearm use (referent no)	173.043	1	4.719	0.031*	0.015
GT score	84.001	1	2.291	0.131	0.007
NEO-FFI					
Neuroticism	11.672	1	0.318	0.573	0.001
Extraversion	5.767	1	0.157	0.692	0.001
Openness to experience	77.751	1	2.120	0.146	0.007
Agreeableness	41.375	1	1.128	0.289	0.004
Conscientiousness	16.079	1	0.438	0.508	0.001
Pittsburgh Sleep Quality Index	38.364	1	1.046	0.307	0.003

*Significant at ≤ 0.05 level.
Notes: GT score = General technical aptitude score; MS = Mean square; NEO-FFI = NEO Five-Factor Inventory.

Table VI-24. Between-participant effects for marksmanship score (continued).

Source	MS	df	*F*	*p*	η^2
Week *t** – 1 Profile of Mood States					
T-factor	86.493	1	2.359	0.126	0.007
D-factor	27.612	1	0.753	0.386	0.002
A-factor	0.089	1	0.002	0.961	0.000
V-factor	0.902	1	0.025	0.876	0.000
F-factor	129.144	1	3.522	0.062	0.011
C-factor	22.697	1	0.619	0.432	0.002
Week *t** Profile of Mood States					
T-factor	0.415	1	0.011	0.915	0.000
D-factor	57.535	1	1.569	0.211	0.005
A-factor	5.613	1	0.153	0.696	0.000
V-factor	46.526	1	1.269	0.261	0.004
F-factor	15.798	1	0.431	0.512	0.001
C-factor	0.325	1	0.009	0.925	0.000
RSES	10.603	1	0.289	0.591	0.001
Sex (referent male)	434.120	1	11.838	0.001*	0.036
Tobacco use (referent no)	7.273	1	0.198	0.656	0.001
Error	36.673	313			

*Significant at ≤ 0.05 level.
Notes: MS = Mean square; RSES = Response to Stressful Experiences Scale.

The analysis was repeated for the subsample of participants for which actigraphy data was available. The same general analytic approach was used except that the average hours slept during the week prior to (t^* – 1) and the week of (t^*) the record fires were used as the covariates. A total of 90 participants, 52 (98% of the initial sub-cohort) in the intervention group and 38 (93% of the initial sub-cohort) in the comparison group, had at least two observations recorded in the marksmanship databases. Table VI-25 displays the results for the within-participant model. Again using a 5% significance level, there was no significant within-participant effect of practice, but there was a significant interaction effect between practice and treatment condition.

Table VI-25. Within-participant effects for marksmanship score for the actigraphy subsample.

Source	MS	df	*F*	*p*	η^2
Practice	5.079	1	0.289	0.593	0.004
Practice x Condition	105.668	1	6.003	0.017*	0.071
Practice x Chronotype	1.681	2	0.095	0.909	0.002
Practice x Condition x Chronotype	3.893	2	0.221	0.802	0.006
Practice x Week *t** – 1 average sleep	65.360	1	3.713	0.058	0.045
Practice x Week *t** average sleep	21.476	1	1.220	0.273	0.015
Error	17.602	78			

*Significant at ≤ 0.05 level.
Note: MS = Mean square.

Although the intervention and comparison groups did not differ in terms of mean initial and final scores, there was a trend for participants in the intervention group to have a greater improvement in scores with practice than participants in the comparison group (Figure VI-37). There was no interaction effect between practice and chronotype, nor were there any interaction effects between practice and the covariates.

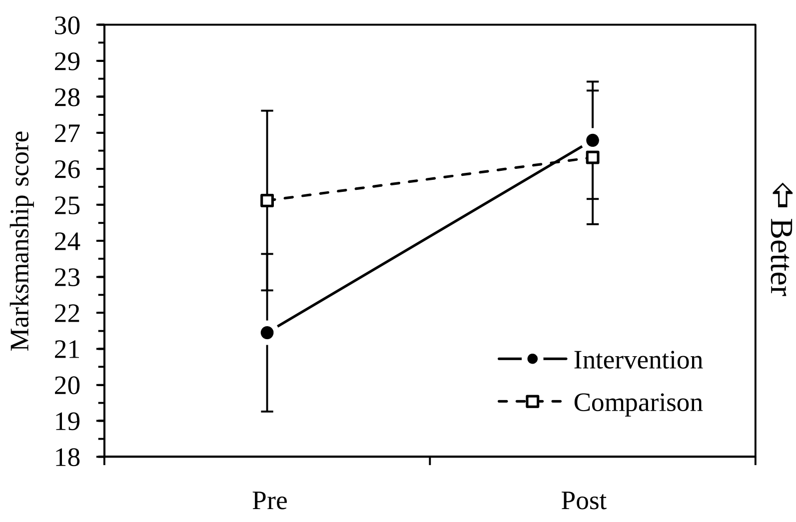


Figure VI-37. Estimated marginal means for first and last marksmanship scores by treatment condition for the actigraphy subsample (error bars are for 95% confidence intervals).

In terms of the between-participant model (Table VI-26), there was no significant fixed effect of treatment condition in the presence of the sleep covariates. Additionally, there was no significant fixed effect for chronotype, nor was there an interaction effect between treatment condition and chronotype. There was, however, a significant effect for the covariate, week $t^* - 1$ average sleep, but not week t^* average sleep.

Table VI-26. Between-participant effects for marksmanship score for the actigraphy subsample.

Source	MS	df	F	p	η^2
Condition	62.723	1	1.439	0.234	0.018
Chronotype	5.237	2	0.120	0.887	0.003
Condition x Chronotype	56.897	2	1.305	0.277	0.032
Week $t^* - 1$ average sleep	177.670	1	4.076	0.047*	0.050
Week t^* average sleep	48.316	1	1.108	0.296	0.014
Error	43.589	78			

*Significant at ≤ 0.05 level.
Note: MS = Mean square.

5. Physical Fitness

It was of interest to determine how participants' performance on the Army Physical Fitness Test related to treatment condition and chronotype while accounting for potential covariates. However, an issue was identified upon receipt of the physical fitness database from each company that needed to be addressed prior to choosing an analytic approach. Forty-nine (12.5%) participants had no scores reported for any of the three physical fitness tests, 10.2% of the remaining 343 participants had no scores reported for either one or two of the physical fitness tests. This issue was addressed by analyzing the physical fitness dataset as a repeated cross-section design rather than a within-participant repeated measures design and using a 1% significance level to counter the resulting increased power of statistical tests.

A MANCOVA of the component physical fitness scores (push-ups, sit-ups, and run) was accomplished using treatment condition, week, and chronotype as fixed effects and age, caffeine and tobacco use, component, fitness factors (BMI and exercise frequency), GT score, personality component scores, RSES score, sex, and sleep factors (ESS and PSQI scores) as covariates. In addition, POMS measurements from the week of the corresponding physical fitness test were also included as covariates. Tables VI-27 and VI-28 summarize the results of the multivariate tests. There were significant fixed effects for treatment condition, week, and chronotype as well as a significant interaction effect between treatment condition and week. There were also significant effects for the covariates age, BMI, exercise frequency, GT score, NEO neuroticism component score, POMS A-factor score, and sex.

Table VI-27. Multivariate tests for physical fitness component scores.

Source	Wilks' λ	F	df1	df2	p	η^2
Condition	0.964	11.037	3	884	<0.001*	0.036
Week	0.955	6.868	6	1768	<0.001*	0.023
Chronotype	0.963	5.676	6	1768	<0.001*	0.019
Condition x Week	0.978	3.317	6	1768	0.003*	0.011
Condition x Chronotype	0.994	0.838	6	1768	0.540	0.003
Chronotype x Week	0.995	0.396	12	2339	0.966	0.002
Condition x Chronotype x Week	0.994	0.425	12	2339	0.954	0.002
Age	0.952	14.765	3	884	<0.001*	0.048
Body mass index	0.887	37.504	3	884	<0.001*	0.113
Caffeine use (referent no)	1.000	0.045	3	884	0.987	<0.001
Component (referent regular)	0.996	1.201	3	884	0.308	0.004
Epworth Sleepiness Scale	0.997	0.919	3	884	0.431	0.003
Exercise frequency	0.981	5.601	3	884	0.001*	0.019
GT score	0.976	7.391	3	884	<0.001*	0.024

*Significant at ≤ 0.01 level.
Notes: GT score = General technical aptitude score.

Table VI-28. Multivariate tests for physical fitness component scores (continued).

Source	Wilks' λ	*F*	df1	df2	*p*	η^2
NEO-FFI						
Neuroticism	0.975	7.442	3	884	<0.001*	0.025
Extraversion	0.990	3.011	3	884	0.029	0.010
Openness to experience	0.999	0.196	3	884	0.899	0.001
Agreeableness	0.999	0.376	3	884	0.770	0.001
Conscientiousness	0.990	2.840	3	884	0.037	0.010
Pittsburgh Sleep Quality Index	0.991	2.758	3	884	0.041	0.009
Profile of Mood States						
T-factor	0.994	1.645	3	884	0.177	0.006
D-factor	0.996	1.120	3	884	0.340	0.004
A-factor	0.976	7.167	3	884	<0.001*	0.024
V-factor	0.995	1.465	3	884	0.223	0.005
F-factor	0.993	2.177	3	884	0.089	0.007
C-factor	0.997	0.918	3	884	0.432	0.003
RSES	0.993	1.948	3	884	0.120	0.007
Sex (referent male)	0.944	17.607	3	884	<0.001*	0.056
Tobacco use (referent no)	0.999	0.242	3	884	0.867	0.001

*Significant at ≤ 0.01 level.
Notes: NEO-FFI = NEO Five-Factor Inventory; RSES = Response to Stressful Experiences Scale.

Table VI-29 provides the results of the relevant univariate tests of between-participant effects for push-up score. There were significant fixed effects for treatment condition and week as well as a significant interaction effect between condition and week. The estimated marginal mean push-up score for the intervention group was 76.404 (99% CI: 73.992, 78.816) versus 70.475 (99% CI: 67.921, 73.028) for the comparison group.

Table VI-29. Univariate tests of between-participant effects for push-up score.

Source	MS	df	*F*	*p*	η^2
Condition	3727.319	1	16.107	<0.001*	0.018
Week	3250.914	2	14.048	<0.001*	0.031
Chronotype	333.852	2	1.443	0.237	0.003
Condition x Week	1588.026	2	6.862	0.001*	0.015
Age	920.453	1	3.978	0.046	0.004
Body mass index	6508.729	1	28.126	<0.001*	0.031
Exercise frequency	3338.788	1	14.428	<0.001*	0.016
GT score	1573.779	1	6.801	0.009*	0.008
NEO-FFI neuroticism	994.902	1	4.299	0.038	0.005
POMS A-factor	842.023	1	3.639	0.057	0.004
Sex (referent male)	1622.487	1	7.011	0.008*	0.008
Error	231.413	886			

*Significant at ≤ 0.01 level.
Notes: GT score = General technical aptitude score; MS = Mean square; NEO-FFI = NEO Five-Factor Inventory; POMS = Profile of Mood States.

For the fixed effect of week (Figure VI-38), the pairwise difference occurred between week 3 versus week 8 ($p < 0.001$). Note that physical fitness assessments were only accomplished on weeks 3, 6, and 9.

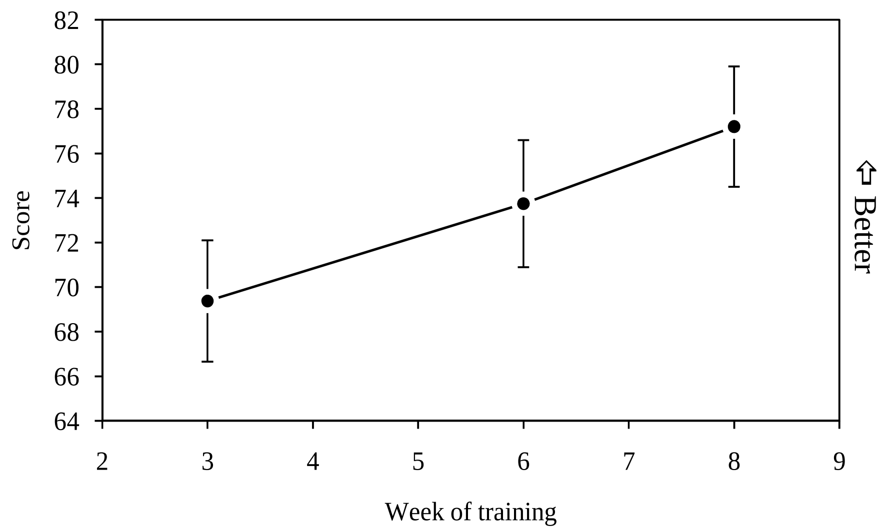


Figure VI-38. Estimated marginal means for push-up score by week of training (error bars are for 99% confidence intervals).

Regarding the interaction effect (Figure VI-39), the intervention and comparison groups differed in mean push-up score at week 3, but participants in the comparison group improved at a faster rate than those in the intervention group such that there were no differences in mean score by weeks 6 and 8.

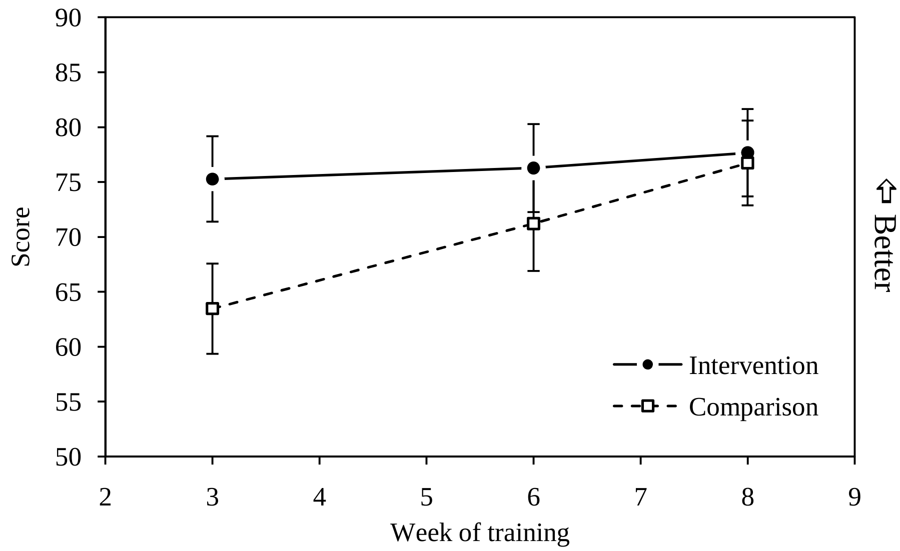


Figure VI-39. Estimated marginal means for push-up score by treatment condition and week of training (error bars are for 99% confidence intervals).

Significant covariates included age, BMI, exercise frequency, GT score, and sex, although BMI and exercise frequency had effect sizes that were two to four times greater than the effect sizes of GT score and sex.

Table VI-30 provides the results of the relevant univariate tests of between-participant effects for sit-up score. There were significant fixed effects for treatment condition, week, and chronotype as well as a significant interaction effect between treatment condition and week. The estimated marginal mean push-up score for the intervention group was 73.128 (99% CI: 70.840, 75.416) versus 68.353 (99% CI: 65.930, 70.775) for the comparison group.

Table VI-30. Univariate tests of between-participant effects for sit-up score.

Source	MS	df	*F*	*p*	η^2
Condition	2417.448	1	11.610	0.001*	0.013
Week	2642.599	2	12.691	<0.001*	0.028
Chronotype	1071.267	2	5.145	0.006*	0.011
Condition x Week	1196.870	2	5.748	0.003*	0.013
Age	159.669	1	0.767	0.381	0.001
Body mass index	9580.624	1	46.010	<0.001*	0.049
Exercise frequency	1782.953	1	8.563	0.004*	0.010
GT score	4162.000	1	19.988	<0.001*	0.022
NEO-FFI neuroticism	2535.853	1	12.178	0.001*	0.014
POMS A-factor	236.754	1	1.137	0.287	0.001
Sex (referent male)	4519.173	1	21.703	<0.001*	0.024
Error	208.227	886			

*Significant at ≤ 0.01 level.

Notes: GT score = General technical aptitude score; MS = Mean square; NEO-FFI = NEO Five-Factor Inventory; POMS = Profile of Mood States.

For the fixed effect of week (Figure VI-40), the pairwise difference occurred between week 3 versus week 8 ($p < 0.001$).

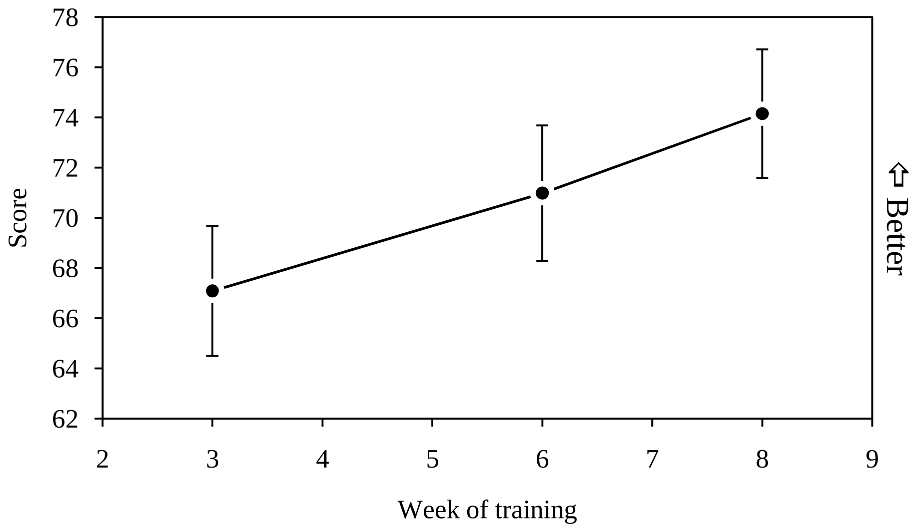


Figure VI-40. Estimated marginal means for sit-up score by week of training (error bars are for 99% confidence intervals).

For the fixed effect of chronotype (Figure VI-41), the pairwise difference occurred between evening versus indeterminate chronotypes (p = 0.004).

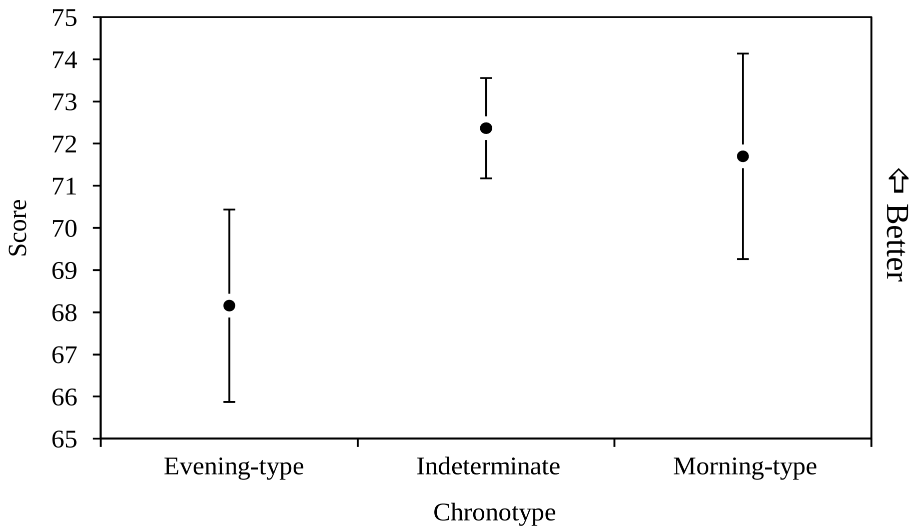


Figure VI-41. Estimated marginal means for sit-up score by chronotype (error bars are for 99% confidence intervals).

Regarding the interaction effect (Figure VI-42), the intervention and comparison groups differed in mean sit-up score at week 3, but participants in the comparison group improved at a faster rate than those in the intervention group such that there were no differences in mean score by weeks 6 and 8.

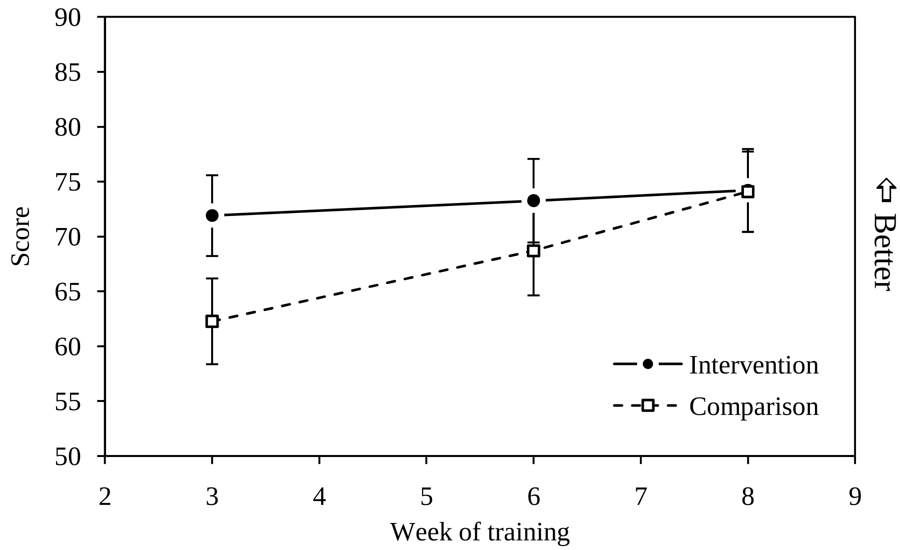


Figure VI-42. Estimated marginal means for sit-up score by treatment condition and week of training (error bars are for 99% confidence intervals).

Significant covariates included BMI, exercise frequency, GT score, NEO neuroticism score, and sex. Body mass index was the most important covariate in terms of effect size.

Table VI-31 provides the results of the relevant univariate tests of between-participant effects for the physical fitness test run score. There was no significant fixed effect for treatment condition, but there were significant fixed effects for week and chronotype.

Table VI-31. Univariate tests of between-participant effects for run score.

Source	MS	df	*F*	*p*	η^2
Condition	435.740	1	1.680	0.195	0.002
Week	2423.699	2	9.346	<0.001*	0.021
Chronotype	3811.444	2	14.697	<0.001*	0.032
Condition x Week	740.598	2	2.856	0.058	0.006
Age	10994.891	1	42.395	<0.001*	0.046
Body mass index	25556.018	1	98.541	<0.001*	0.100
Exercise frequency	354.690	1	1.368	0.243	0.002
GT score	2126.456	1	8.199	0.004*	0.009
NEO-FFI neuroticism	565.387	1	2.180	0.140	0.002
POMS A-factor	5532.681	1	21.333	<0.001*	0.024
Sex (referent male)	367.816	1	1.418	0.234	0.002
Error	259.343	886			

*Significant at ≤ 0.01 level.
Notes: GT score = General technical aptitude score; MS = Mean square; NEO-FFI = NEO Five-Factor Inventory; POMS = Profile of Mood States.

For the fixed effect of week (Figure VI-43), pairwise differences occurred between week 3 versus both week 6 and week 8 ($p < 0.002$).

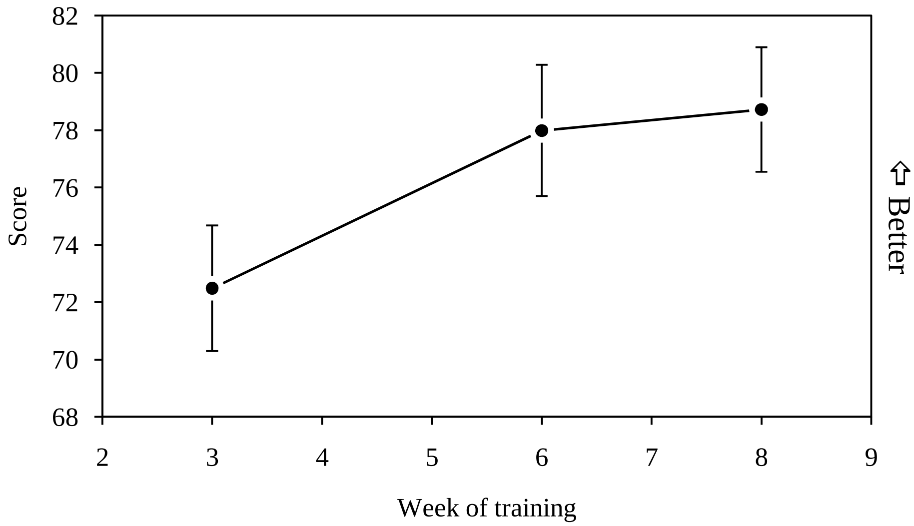


Figure VI-43. Estimated marginal means for run score by week of training (error bars are for 99% confidence intervals).

For the fixed effect of chronotype (Figure VI-44), pairwise differences occurred between evening versus both indeterminate and morning chronotypes ($p < 0.009$). Thus, evening chronotypes were slower than indeterminate and morning chronotypes.

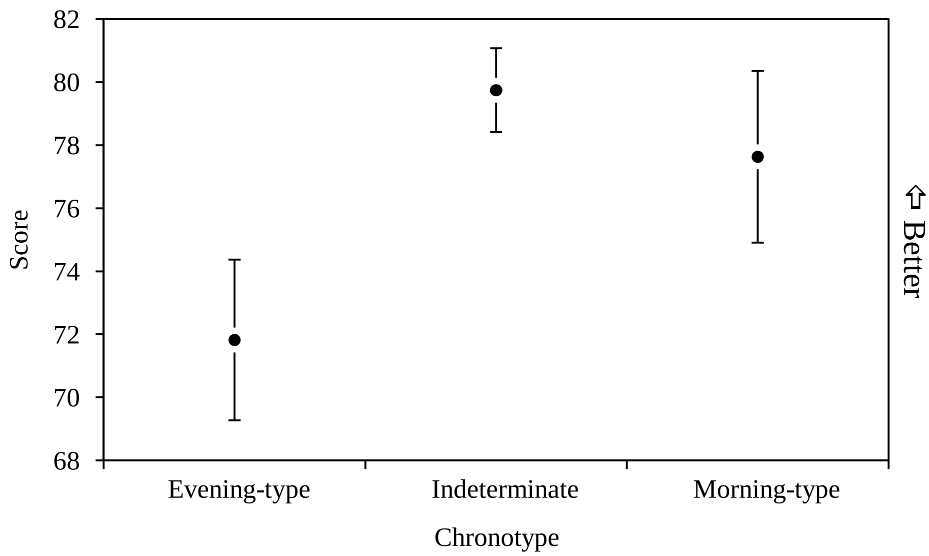


Figure VI-44. Estimated marginal means for run score by chronotype (error bars are for 99% confidence intervals).

There was no significant interaction effect between study condition and week. Significant covariates included age, BMI, GT score, and POMS A-factor score, although BMI was the most important covariate in terms of effect size.

The Army Physical Fitness Test (APFT) score provides a single global estimate of physical fitness and is obtained by summing the scores across the three component fitness assessment activities. An ANCOVA of APFT score was accomplished using treatment condition, week, and chronotype as fixed effects and age, caffeine and tobacco use, component, fitness factors (BMI and exercise frequency), GT score, personality component scores, POMS factor scores, RSES score, sex, and sleep factors (ESS and PSQI scores) as covariates (Tables VI-32 and VI-33). There was no significant fixed effect for treatment condition, but there were significant fixed effects for week and chronotype as well as a significant interaction effect between treatment condition and week.

Table VI-32. Univariate tests of between-participant effects for Army Physical Fitness Test score.

Source	MS	df	*F*	*p*	η^2
Condition	7867.295	1	6.214	0.013	0.007
Week	24182.956	2	19.102	<0.001*	0.041
Chronotype	12473.396	2	9.853	<0.001*	0.022
Condition x Week	9496.913	2	7.501	0.001*	0.017
Condition x Chronotype	453.751	2	0.358	0.699	0.001
Chronotype x Week	779.760	4	0.616	0.651	0.003
Condition x Chronotype x Week	752.311	4	0.594	0.667	0.003
Age	21989.056	1	17.369	<0.001*	0.019
Body mass index	114926.602	1	90.779	<0.001*	0.093
Caffeine use (referent no)	20.595	1	0.016	0.899	<0.001
Component (referent regular)	32.099	1	0.025	0.874	<0.001
Epworth Sleepiness Scale	3086.853	1	2.438	0.119	0.003
Exercise frequency	14194.166	1	11.212	0.001*	0.012
GT score	23105.988	1	18.251	<0.001*	0.020

*Significant at ≤ 0.01 level.
Notes: GT score = General technical aptitude score; MS = Mean square.

Table VI-33. Univariate tests of between-participant effects for Army Physical Fitness Test score (continued).

Source	MS	df	*F*	*p*	η^2
NEO-FFI					
Neuroticism	3257.315	1	2.573	0.109	0.003
Extraversion	2419.963	1	1.911	0.167	0.002
Openness to experience	335.026	1	0.265	0.607	<0.001
Agreeableness	949.270	1	0.750	0.387	0.001
Conscientiousness	192.961	1	0.152	0.696	<0.001
Profile of Mood States					
T-factor	5577.076	1	4.405	0.036	0.005
D-factor	81.731	1	0.065	0.799	<0.001
A-factor	14252.349	1	11.258	0.001*	0.013
V-factor	5049.279	1	3.988	0.046	0.004
F-factor	3378.278	1	2.668	0.103	0.003
C-factor	280.387	1	0.221	0.638	<0.001
RSES	3535.514	1	2.793	0.095	0.003
Sex (referent male)	2184.334	1	1.725	0.189	0.002
Tobacco use	179.788	1	0.142	0.706	<0.001
Error	1266.002	886			

*Significant at ≤ 0.01 level.

Notes: MS = Mean square; NEO-FFI = NEO Five-Factor Inventory; RSES = Response to Stressful Experiences Scale.

For the fixed effect of week (Figure VI-45), pairwise differences in APFT scores occurred between week 3 versus both week 6 and week 8 ($p < 0.001$).

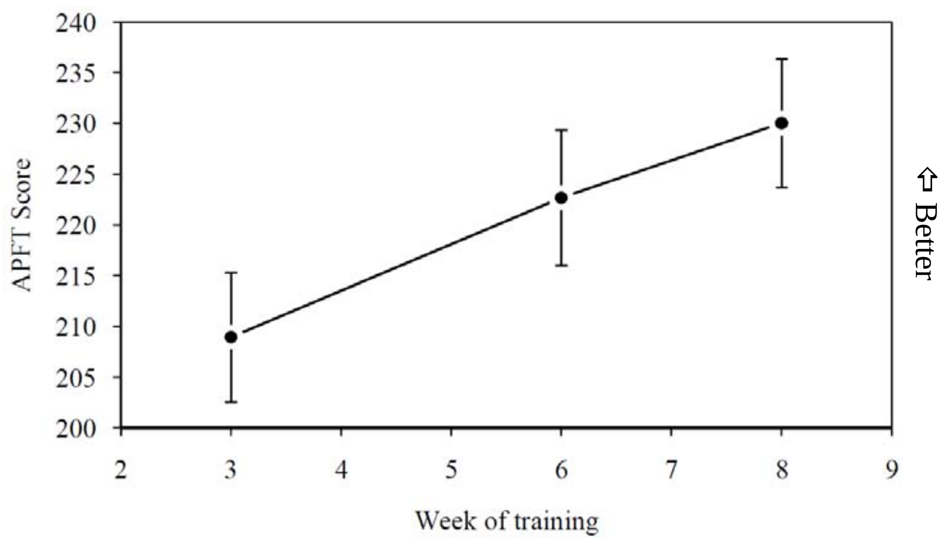


Figure VI-45. Estimated marginal means for Army Physical Fitness Test (APFT) score by week of training (error bars are for 99% confidence intervals).

For the fixed effect of chronotype (Figure VI-46), the pairwise difference in APFT scores occurred between evening versus indeterminate chronotypes ($p < 0.001$).

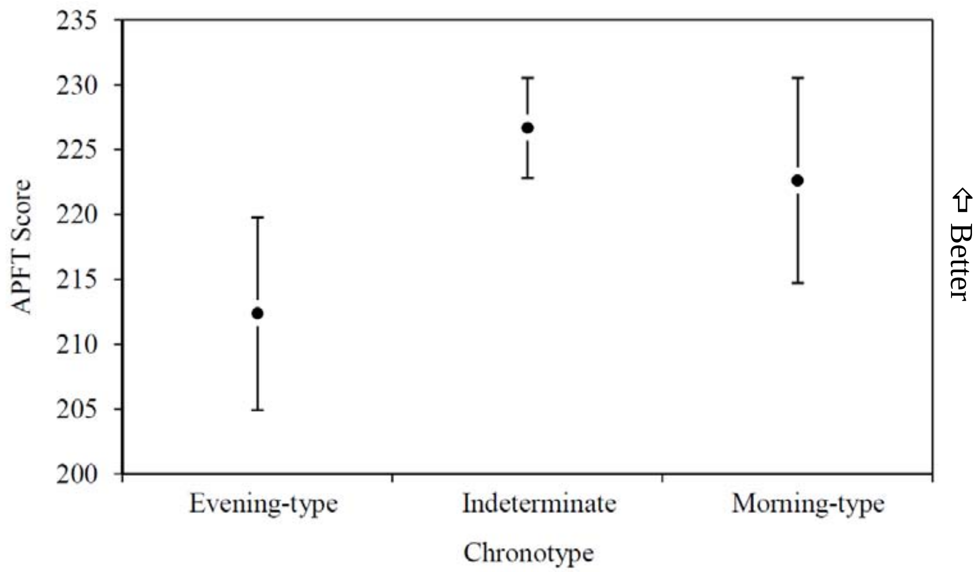


Figure VI-46. Estimated marginal means for Army Physical Fitness Test (APFT) score by chronotype (error bars are for 99% confidence intervals).

Regarding the interaction effect (Figure VI-47), the intervention and comparison groups differed in mean APFT score at week 3, but participants in the comparison group improved at a faster rate than those in the intervention group such that there were no differences in mean score by weeks 6 and 8.

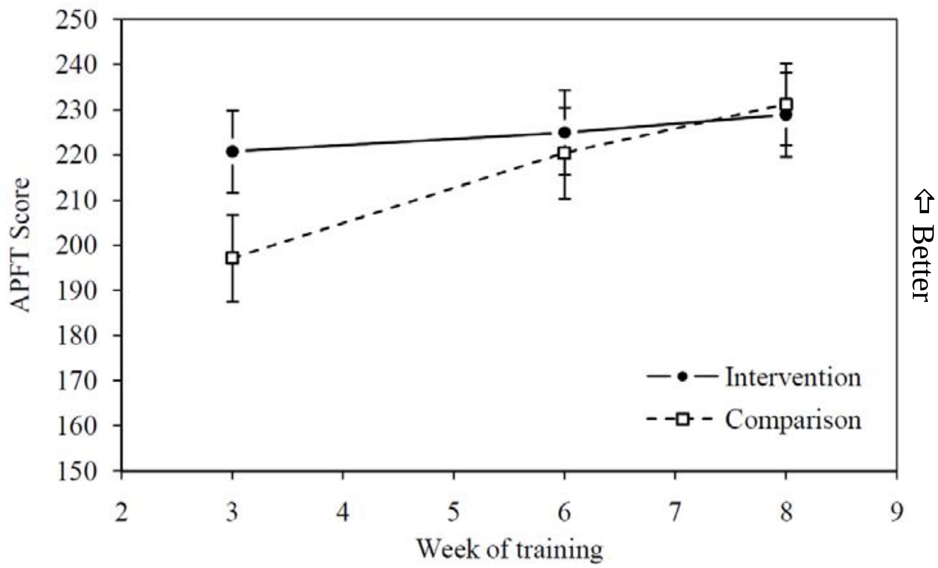


Figure VI-47. Estimated marginal means for Army Physical Fitness Test (APFT) score by treatment condition and week of training (error bars are for 99% confidence intervals).

Significant covariates included age, BMI, exercise frequency, GT score, and POMS A-factor score, but BMI was clearly the most important covariate based on effect size. The analysis of the fitness data was repeated for the subsample of participants for which actigraphy data was available. The same analytic approach was used with the exception that average hours slept per week was used as the covariate. Multivariate tests showed that there was not a significant overall effect of average hours slept per week. Similarly, the univariate analysis of APFT scores for the subsample of participants with actigraphy data showed no significant effect for the covariate, average hours slept per week.

6. Post-Study Questionnaire

Both the pre-study and post-study questionnaires assessed participant sleep using two standardized survey instruments: the Epworth Sleepiness Scale (ESS) and the Pittsburgh Sleep Quality Index (PSQI). The effect of the treatment intervention on ESS and PSQI scores was assessed using a pre/post study design. A repeated measures ANCOVA of ESS and PSQI scores was accomplished using time as a within-participant effect; treatment condition and chronotype as fixed between-participant effects; and age, caffeine and tobacco use, component, firearm use, fitness factors (BMI and exercise frequency), GT score, personality component scores, RSES score, and sex as covariates. Because of participant attrition, there were missing post-study questionnaires for 44 participants (21%) in the intervention group and 31 participants (17%) in the comparison group. This difference was not statistically significant.

a. Epworth Sleepiness Scale

Based on a 5% significance level, in terms of within-participant effects of Epworth Sleeping Scale (ESS) score (Table VI-34), there was no significant within-participant effect of time, nor was there a significant interaction effect between time and chronotype. There were significant interaction effects between time and the fixed effect, treatment condition, as well as the covariate GT score.

Table VI-34. Within-participant effects for Epworth Sleepiness Scale score.

Source	MS	df	*F*	*p*	η^2
Time	3.157	1	0.231	0.631	0.001
Time x Condition	259.141	1	18.943	<0.001*	0.060
Time x Chronotype	7.304	2	0.534	0.587	0.004
Time x Condition x Chronotype	14.891	2	1.089	0.338	0.007
Time x Age	2.853	1	0.209	0.648	0.001
Time x Body mass index	7.710	1	0.564	0.453	0.002
Time x Caffeine use (referent no)	1.979	1	0.145	0.704	<0.001
Time x Component (referent regular)	0.406	1	0.030	0.863	<0.001
Time x Exercise frequency	4.765	1	0.348	0.556	0.001
Time x Firearm use (referent no)	13.056	1	0.954	0.329	0.003
Time x GT score	111.942	1	8.183	0.005*	0.027
Time x NEO neuroticism	0.476	1	0.035	0.852	<0.001
Time x NEO extraversion	0.261	1	0.019	0.890	<0.001
Time x NEO openness to experience	4.235	1	0.310	0.578	0.001
Time x NEO agreeableness	44.847	1	3.278	0.071	0.011
Time x NEO conscientiousness	4.997	1	0.365	0.546	0.001
Time x RSES	0.091	1	0.007	0.935	<0.001
Time x Sex (referent male)	38.794	1	2.836	0.093	0.009
Time x Tobacco (referent no)	3.389	1	0.248	0.619	0.001
Error	13.680	296			

*Significant at ≤ 0.05 level.
Notes: MS = Mean square; RSES = Response to Stressful Experiences Scale.

The interaction effect between time and treatment condition is shown in Figure VI-48. ESS scores increased significantly for participants in the comparison group over the course of training but remained unchanged for those in the intervention group. Consequently, the groups' mean scores differed significantly at the post-study assessment with the comparison group reporting greater sleepiness.

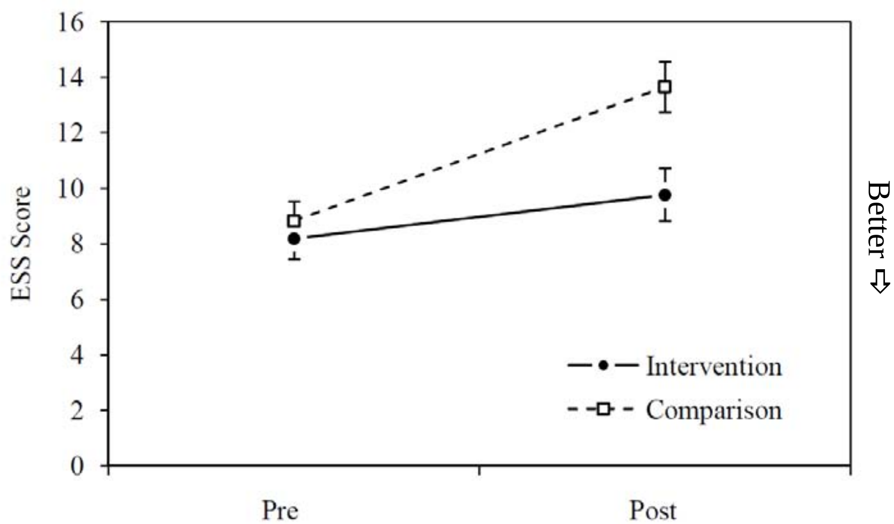


Figure VI-48. Estimated marginal means for ESS score by treatment condition and week of training (error bars are for 95% confidence intervals).

In terms of between-participant effects for ESS score (Table VI-35), there was a significant fixed effect of treatment condition, with an estimated marginal mean ESS score of 8.978 (95% CI: 8.297, 9.659) in the intervention group versus 11.242 (95% CI: 10.595, 11.888) in the comparison group.

Table VI-35. Between-participant effects for Epworth Sleepiness Scale score.

Source	MS	df	*F*	*p*	η^2
Condition	503.762	1	21.635	<0.001*	0.068
Chronotype	104.965	2	4.508	0.012*	0.030
Condition x Chronotype	3.886	2	0.167	0.846	0.001
Age	0.156	1	0.007	0.935	<0.001
Body mass index	4.916	1	0.211	0.646	0.001
Caffeine use (referent no)	5.897	1	0.253	0.615	0.001
Component (referent regular)	20.799	1	0.893	0.345	0.003
Exercise frequency	14.138	1	0.607	0.436	0.002
Firearm use (referent no)	17.778	1	0.764	0.383	0.003
GT score	70.499	1	3.028	0.083	0.010
NEO-FFI					
Neuroticism	27.178	1	1.167	0.281	0.004
Extraversion	34.900	1	1.499	0.222	0.005
Openness to experience	29.898	1	1.284	0.258	0.004
Agreeableness	13.613	1	0.585	0.445	0.002
Conscientiousness	12.016	1	0.516	0.473	0.002
RSES	49.023	1	2.105	0.148	0.007
Sex (referent male)	345.942	1	14.857	<0.001*	0.048
Tobacco use	96.270	1	4.135	0.043*	0.014
Error	23.285	296			

*Significant at ≤ 0.05 level.
Notes: GT score = General technical aptitude score; MS = Mean square; NEO-FFI = NEO Five-Factor Inventory; RSES = Response to Stressful Experiences Scale.

There was also a significant fixed effect of chronotype (Figure VI-49), with the pairwise difference in ESS score occurring between evening and morning chronotypes (p = 0.009). There was no significant interaction effect for ESS score between treatment condition and chronotype. Significant covariates included sex and tobacco use with females and smokers reporting greater sleepiness.

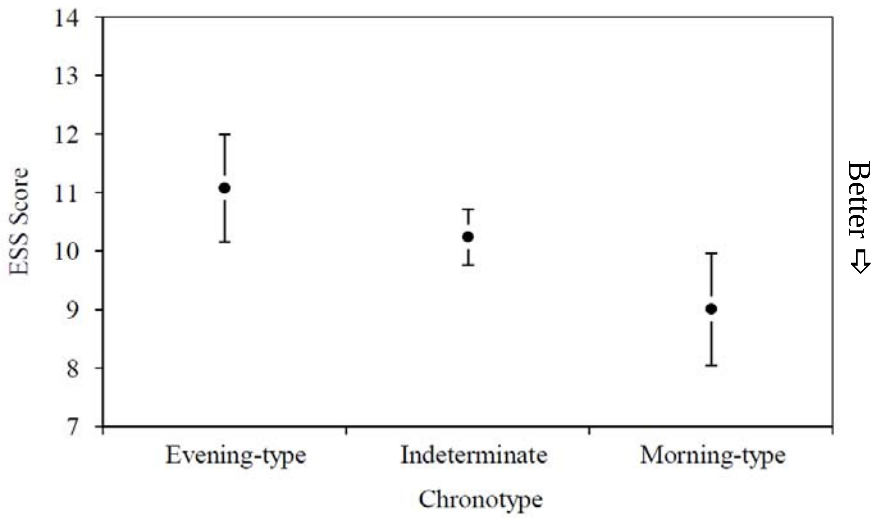


Figure VI-49. Estimated marginal means for ESS score by chronotype (error bars are for 95% confidence intervals).

Scores above ten on the ESS are indicative of excessive sleepiness and are a cause for concern with respect to performance (Miller, 2006). Applying this standard to our study sample, the odds ratio for a participant reporting excessive sleepiness being in the comparison relative to the intervention group was 1.198 (95% CI: 0.765, 1.874) prior to training and 2.331 (95% CI: 1.478, 3.679) at the completion of training. There was no difference in the odds of participants in the intervention and comparison groups being excessively sleepy at the start of training. However, participants in the comparison group were approximately 1.5 to 3.5 times more likely to be excessively sleepy by the conclusion of training, indicative of their sleep debt accrual throughout the course of Basic Combat Training.

b. Pittsburgh Sleep Quality Index

In terms of within-participant effects of Pittsburgh Sleep Quality Index (PSQI) score (Table VI-36), there was no significant fixed effect of time, nor was there a significant interaction effect between time and chronotype.

Table VI-36. Within-participant effects for Pittsburgh Sleep Quality Index score.

Source	MS	df	*F*	*p*	η^2
Time	0.297	1	0.044	0.834	<0.001
Time x Condition	163.180	1	24.125	<0.001*	0.075
Time x Chronotype	15.529	2	2.296	0.102	0.015
Time x Condition x Chronotype	16.370	2	2.420	0.091	0.016
Time x Age	28.914	1	4.275	0.040*	0.014
Time x Body mass index	0.180	1	0.027	0.870	<0.001
Time x Caffeine use (referent no)	0.015	1	0.002	0.962	<0.001
Time x Component (referent regular)	0.046	1	0.007	0.934	<0.001
Time x Exercise frequency	12.623	1	1.866	0.173	0.006
Time x Firearm use (referent no)	1.433	1	0.212	0.646	0.001
Time x GT score	1.170	1	0.173	0.678	0.001
Time x NEO neuroticism	6.520	1	0.964	0.327	0.003
Time x NEO extraversion	0.758	1	0.112	0.738	<0.001
Time x NEO openness to experience	6.487	1	0.959	0.328	0.003
Time x NEO agreeableness	3.250	1	0.481	0.489	0.002
Time x NEO conscientiousness	9.862	1	1.458	0.228	0.005
Time x RSES	0.526	1	0.078	0.781	<0.001
Time x Sex (referent male)	0.048	1	0.007	0.933	<0.001
Time x Tobacco (referent no)	0.215	1	0.032	0.859	<0.001
Error	6.764	296			

*Significant at ≤ 0.05 level.
Notes: GT score = General technical aptitude score; MS = Mean square; RSES = Response to Stressful Experiences Scale.

There were significant interaction effects of PSQI score between time and the fixed effect, treatment condition, as well as the covariate age. The interaction effect with treatment condition is shown in Figure VI-50. PSQI scores increased for participants in the comparison group and decreased for participants in the intervention group over the course of training such that the groups mean scores differed significantly at the post-study assessment.

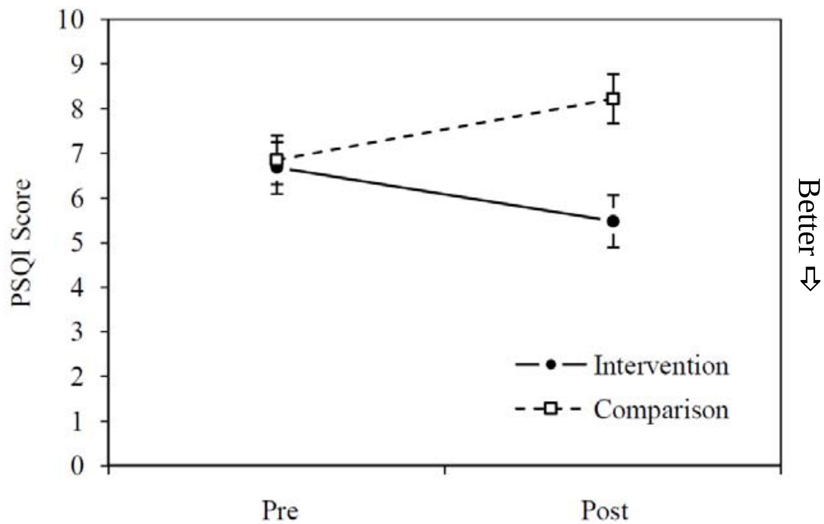


Figure VI-50. Estimated marginal means for PSQI score by treatment condition and pre/post-training (error bars are for 95% confidence intervals).

In terms of between-participant effects of PSQI score (Table VI-37), there was a significant fixed effect of treatment condition, with an estimated marginal mean PSQI score of 6.082 (95% CI: 5.629, 6.536) in the intervention group versus 7.539 (95% CI: 7.109, 7.970) in the comparison group. There was no significant fixed effect of chronotype, nor was there a significant interaction effect between treatment condition and chronotype. Significant covariates included age and the NEO personality components of neuroticism, openness to experience, agreeableness, and conscientiousness scores.

Table VI-37. Between-participant effects for Pittsburgh Sleep Quality Index score.

Source	MS	df	*F*	*p*	η^2
Condition	208.769	1	20.244	<0.001*	0.064
Chronotype	9.839	2	0.954	0.386	0.006
Condition x Chronotype	9.636	2	0.934	0.394	0.006
Age	185.963	1	18.033	<0.001*	0.057
Body mass index	17.835	1	1.729	0.189	0.006
Caffeine use (referent no)	8.543	1	0.828	0.363	0.003
Component (referent regular)	1.432	1	0.139	0.710	<0.001
Exercise frequency	19.454	1	1.886	0.171	0.006
Firearm use (referent no)	30.064	1	2.915	0.089	0.010
GT score	33.465	1	3.245	0.073	0.011
NEO-FFI					
Neuroticism	97.425	1	9.447	0.002*	0.031
Extraversion	2.788	1	0.270	0.603	0.001
Openness to experience	89.635	1	8.692	0.003*	0.029
Agreeableness	180.261	1	17.480	<0.001*	0.056
Conscientiousness	47.638	1	4.619	0.032*	0.015
RSES	5.616	1	0.545	0.461	0.002
Sex (referent male)	17.329	1	1.680	0.196	0.006
Tobacco use	3.049	1	0.296	0.587	0.001
Error	10.312	296			

*Significant at ≤ 0.05 level.
Notes: GT score = General technical aptitude score; MS = Mean square; NEO-FFI = NEO Five-Factor Inventory; RSES = Response to Stressful Experiences Scale.

Scores above five on the PSQI are indicative of poor sleep quality. Applying this standard to our study sample, the odds ratio for a participant having poor quality sleep being in the comparison relative to the intervention group was 1.684 (95% CI: 1.106, 2.565) prior to training and 5.477 (95% CI: 3.343, 8.972) at the completion of training. Moreover, the odds of a participant having poor sleep quality decreased in the intervention group from pre-training (odds = 0.791; 95% CI: 0.659, 0.950) to post-training (odds = 0.470; 95% CI: 0.377, 0.586). In contrast, the odds of a participant

having poor sleep quality increased in the comparison group from pre-training (odds = 1.332; 95% CI: 1.047, 1.696) to post-training (odds = 2.574; 95% CI: 1.889, 2.509).

c. Ordinal Sleep Ratings

Participants provided ordinal ratings of the adequacy of the sleep obtained by themselves and peers using a 5-item Likert scale. Figure VI-51 provides histograms of the participants' ratings by treatment condition.

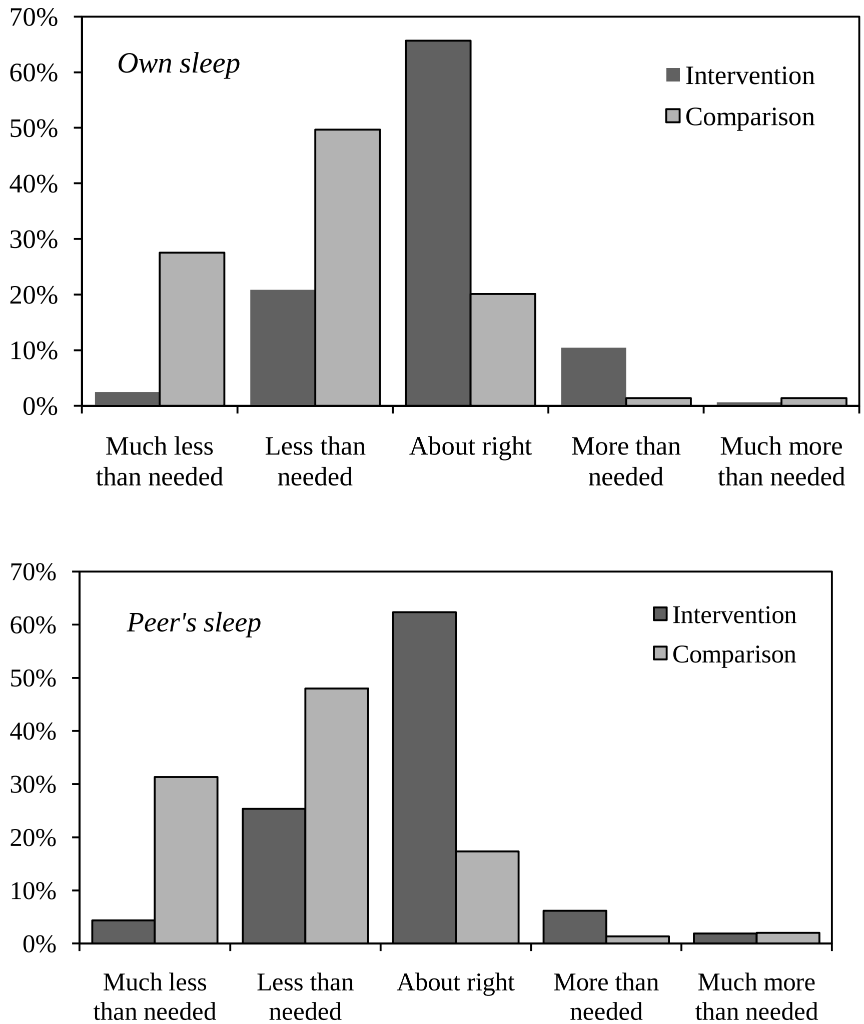


Figure VI-51. Histogram of participants' ratings of their own sleep (top) and their peers' sleep (bottom) by treatment condition.

The distributions of ratings by the comparison group were positively skewed versus those of the intervention group, which were symmetric unimodal. The mean rank for both ratings was higher for the intervention group than the comparison group: own sleep (intervention mean rank = 203.0, comparison mean rank = 110.5, Mann-Whitney U = 5164.5, $p < 0.001$) and peers' sleep (intervention mean rank = 198.6, comparison mean rank = 112.4, U = 5495.0, $p < 0.001$). There were small to moderate negative correlations between participants' ordinal ratings of the adequacy of their own sleep and their post-training ESS ($\rho = -0.351$, $p < 0.001$) and PSQI scores ($\rho = -0.505$, $p < 0.001$). Similarly, there was a negative correlation between participants' own sleep ratings and post-training POMS total mood disturbance scores ($\rho = -0.370$, $p < 0.001$).

d. Frequency of Sleep During Activities

Participants were asked to report, on average, how often they fell asleep during activities such as classes, training, or lectures. Figure VI-52 provides a histogram of the participants' responses by treatment condition.

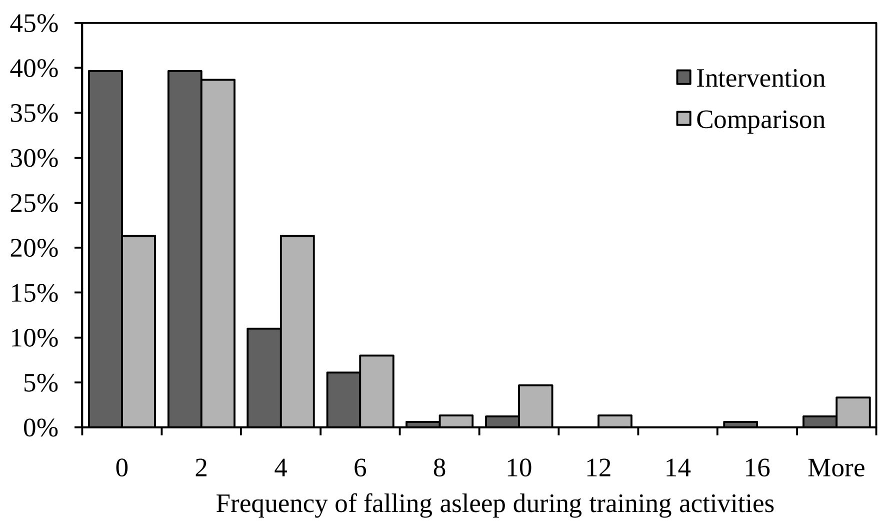


Figure VI-52. Histogram of daily frequency that participants report falling asleep during activities by treatment condition.

The distribution of responses for both groups was positively skewed, but the distribution for the intervention group was platykurtic at 0–2 while that of the comparison group was mesokurtic between 0–4. A comparison of mean ranks confirmed that participants in the intervention group reported significantly fewer episodes of falling asleep than those in the comparison group (intervention mean rank = 137.5, comparison mean rank = 179.4, Mann-Whitney U = 9011.0, $p < 0.001$). There was a small positive correlation between the frequency that participants fell asleep during activities and their post-training ESS ($\rho = 0.365$, $p < 0.001$) and PSQI scores ($\rho = 0.291$, $p < 0.001$). There was also a positive correlation between the frequency that participants fell asleep during activities the post-training POMS total mood disturbance score ($\rho = 0.206$, $p < 0.001$). Additionally, there was a small negative correlation between a participant's ordinal rating of their sleep and the frequency with which they reported falling asleep during activities ($\rho = -0.250$, $p < 0.001$).

e. Preference in Timing of Physical Fitness Training

Participants were asked to indicate their preference for the best time of day for physical fitness training. Figure VI-53 provides a histogram of the participants' responses by treatment condition.

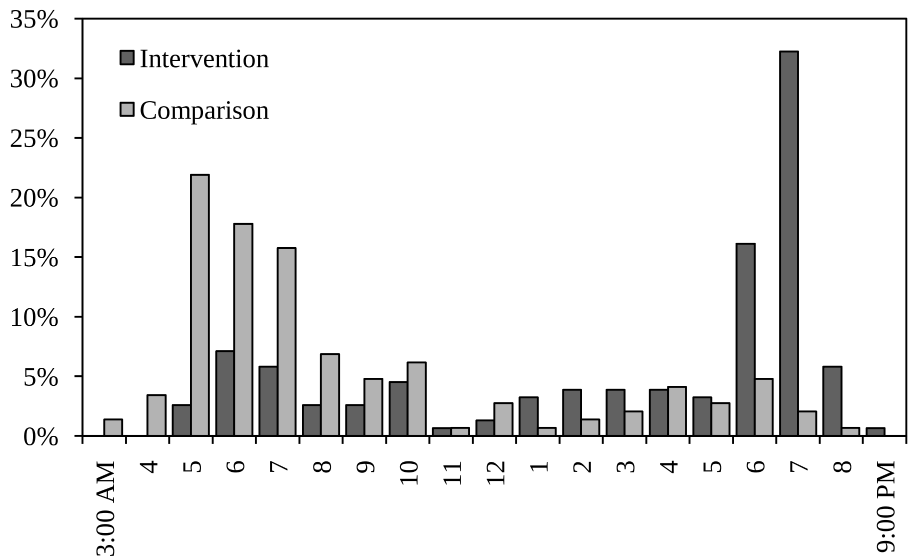


Figure VI-53. Histogram of participants' preferred time of the day for physical fitness training by treatment condition.

The distribution of responses for both groups was bimodal with the primary peak occurring near the respectively scheduled company physical fitness training times. Hence, participants in the intervention group indicated they generally preferred to conduct physical fitness training in the evenings as per their training schedule. Similarly, participants in the comparison group preferred to conduct physical fitness training in the mornings as per their training schedule. There was a small, negative correlation between a participant's Morningness-Eveningness Questionnaire score and their time preference for physical fitness training ($\rho = -0.272$, $p < 0.001$). Thus, evening chronotype participants preferred physical fitness training in the evening and morning chronotype participants preferred training in the morning.

7. Attrition

It was of interest to determine how participants' likelihood of completing training related to treatment condition and other potential measured covariates. The databases submitted by each of the training companies indicated whether each participant successfully completed training. However, for those participants who did not complete training, the databases did not uniformly indicate when an attrition occurred and for what reason. Moreover, the final disposition of participants who did not graduate was not always determined, with some being separated from the Army, others on convalescent leave pending recovery from an injury or awaiting a physical evaluation board, and still others washing back to reaccomplish either portions of or the entire course of training. Additionally, participants who did not meet physical fitness standards could also be sent to a special training company to focus on further physical conditioning. Thus, a participant being classified as an attrite does not necessarily equate with them being lost to the Army. Accordingly, it was decided to analyze the likelihood of a participant not graduating with their initial training cohort using a simple binary logistic regression model and limiting the covariates to those measured during the initial study enrollment.

Overall, 35 (16.7%) participants in the intervention group failed to graduate with their cohort as compared to 33 (18.1%) participants in the comparison group, a non-significant difference $\left(\chi_1^2 = 0.130, p = 0.718\right)$. Table VI-38 shows the results for the

fitted binary logistic regression model for failure to graduate. Accordingly, the odds ratios (ORs), calculated from the exponential of the estimated regression coefficients, should be interpreted in terms of the likelihood of failing to graduate with one's initial training cohort. The classification accuracy of the model was 83.9% using a cutoff of 0.5.

Table VI-38. Results for the fitted binary logistic regression model for failure to graduate with initial training cohort.

Analysis variables	Estimate	Standard error	df	Wald	*p*
Intercept	-7.387	1.317	1	31.450	<0.001
Body mass index	0.104	0.033	1	9.882	0.002
NEO-FFI neuroticism	0.039	0.017	1	5.507	0.019
POMS depression-dejection factor	0.024	0.011	1	4.651	0.031
Sex (referent male)	1.514	0.314	1	23.236	<0.001

There was no significant effect of treatment condition on the likelihood of failure to graduate. However, being female (OR = 4.545; 95% CI: 2.456, 8.411), increased body mass index (OR = 1.110; 95% CI: 1.1040, 1.184), higher scores of neuroticism as assessed using the NEO-FFI (OR = 1.040; 95% CI: 1.006, 1.074), and depressed mood or sense of inadequacy as measured on the POMS (OR = 1.024; 95% CI: 1.002, 1.046) were all associated with an increased likelihood of failure to graduate.

E. DISCUSSION

Most studies of training effectiveness in military environments have concerned themselves primarily with activities that occur during the waking hours. They tend to examine the relationship between time expenditures in training using various modalities and measures of individual or system performance—the archetype being the classic transfer of training study. This study took a decidedly different approach, instead concerning itself primarily with the importance of the hours spent sleeping and their relation to measures of Soldier performance and other indicators of individual functioning during basic combat training. Recognizing that adolescents comprise the

majority of military accessions, this study evaluated the impact of accommodating adolescent alterations in sleeping and waking patterns. In particular, the scheduled timing of sleep during training was adjusted to account for the developmental phase delay of the circadian cycle in adolescents. The results of this study indicate that, even after controlling for factors contributing to individual differences, adjusting the scheduled sleep period in a phase delayed direction was associated with increased daily total sleep and modest improvements in some indicators of daytime functioning. These findings suggest several operationally-relevant effects of accommodating adolescent sleep physiology that military planners may wish to consider in developing future training programs of instruction and associated training schedules.

1. Actigraphic Measures of Sleep

Hypothesis 1 predicted that participants on the modified, phase-delayed sleep schedule would obtain more daily sleep than participants following the standard Basic Combat Training schedule. This hypothesis was supported with participants on the modified sleep schedule obtaining approximately 33 more minutes of total sleep per night than those on the standard sleep schedule. This finding is consistent with that of other studies, such as the School Transition Study (Carskadon, 2001), which have found that early start times are associated with truncated sleep in adolescents. The observed reduction in sleep with early start times is attributed to the developmental phase delay of the circadian cycle in adolescents, which makes it particularly difficult for adolescents to advance the evening retiring time in order to obtain an adequate amount of sleep. Additionally, Carskadon and colleagues (1998) have demonstrated that adolescents do not readily adapt or habituate their circadian cycle to early rising times, although the mechanism underlying this observation is not well understood. It is also interesting to note that a similar phenomenon has been described in adult shift workers with very early morning starts who tend to experience long sleep latencies when attempting to get compensatory sleep in the early evening (Rosa, 2001).

Thus, this study demonstrates that scheduling the sleep period for adolescents and young adults to better align with the phase delay in their circadian cycle results in a

significant improvement in total daily sleep without any concomitant adjustment to the quantity of time scheduled for sleep. Regardless of differences in the timing of sleep between the two schedules, morning chronotype participants averaged approximately 15 minutes more sleep than those participants who were evening chronotype. This pattern is consistent with that described by Wolfson (2001) for adolescent students transitioning to a school with an earlier start time: evening chronotype students had more difficulty adjusting to the earlier start time and had less total sleep than did morning chronotype students. The implication is that even with the phase-delayed schedule used in this study, evening chronotype participants experienced greater difficulty adjusting to their new start time. This result is not surprising given the histograms of participants' self-reported wake times prior to Basic Combat Training, which suggest that the transition to military life necessitated earlier start times for the majority of participants. It is also worth noting that the average quantity of sleep obtained by participants was only approximately 60% of the 9.2 hours of daily sleep reportedly needed by adolescents (Mercer, Merritt, & Cowell, 1998; Wolfson, 2001). Lastly, the observation that sleep was reduced for participants using the modified schedule after the sixth week of training is an artifact caused by the commencement of the field exercise portion of Basic Combat Training.

2. Mood States

Hypothesis 2 predicted that participants on the modified sleep schedule would have less decrement in mood state than participants following the standard Basic Combat Training sleep schedule. There was weak support for this hypothesis based on the analysis of the entire study sample, which necessarily excluded consideration of a total daily sleep variable in the models. Irrespective of treatment condition, the general trend was for participants to report decreased feelings of tension-anxiety, depression-dejection, fatigue-inertia, and confusion-bewilderment over the course of Basic Combat Training. Participants in the intervention group reported more stable feelings of anger-hostility and exhibited steadier total mood disturbance scores than participants in the comparison group. Participants in the intervention group also tended towards less anger-hostility and lower total mood disturbance scores relative to the comparison group early in training, although these differences declined during Basic Combat Training. Participants in the

intervention group reported significantly greater feelings of vigor than those in the comparison group throughout training, but the effect size of treatment condition was very modest in this case. Overall, there was no evidence that characteristics of chronotype significantly affected participants' mood states.

There was partial support for Hypothesis 2, particularly with regards to the effects for the characteristics of chronotype on mood, when the analysis was restricted to the actigraphy subsample and a variable for total daily sleep was included in the models. Irrespective of treatment condition, evening chronotype participants reported more vigor throughout training than morning chronotype participants. However, evening chronotype participants in the intervention group exhibited less self-reported feelings of tension-anxiety, depression-dejection, anger-hostility, and confusion-bewilderment than their morning chronotype counterparts. The opposite pattern occurred in the comparison group, with evening chronotype participants reporting greater feelings of tension-anxiety, depression-dejection, anger-hostility, and confusion-bewilderment than their evening chronotype counterparts. In terms of total mood disturbance score, evening chronotype participants in the intervention group had lower scores than their morning chronotype counterparts, while a trend in the opposing direction was observed for participants in the comparison group. Taken together, these findings suggest that the phase-delayed sleep schedule preferentially impacted, in a positive direction, the mood state of evening chronotype participants. The operational significance of this finding is evident when one appreciates that the majority of military accessions are adolescents who, as a demographic group, tend to exhibit a biological predisposition for eveningness (Carskadon, 2001).

The rather modest impact of the sleep schedule intervention on subjective mood in this study contrasts with other research that has shown that manipulations of the duration and timing of sleep episodes can have marked impacts on mood (Birchler-Pedross et al., 2009; Boivin et al., 1997; Danilenko, Cajochen, & Wirz-Justice, 2003; Monk et al., 1992; Selvi et al., 2007; Taub & Berger, 1974; Wood & Magnello, 1992). For example, Boivin and colleagues (1997) demonstrated that even moderate changes in the timing of the sleep-wake cycle led to profound effects on mood. Similarly, Danilenko

and colleagues (2003) showed that advancing the sleep-wake cycle daily by just 20 minutes for a week led to significant decrements in subjective mood ratings relative to a control group with stable sleep. Interestingly, Selvi and colleagues (2007) showed that phase preference modified the effect of partial sleep deprivation on mood, with morning chronotypes exhibiting less sensitivity of mood. A pattern similar to that described by Selvi and colleagues was observed, at least for the subsample of the study population who had actigraphy data.

Several hypotheses are suggested to explain the small observed effect of the schedule intervention on subjective mood in this study. Mood is largely a function of situational factors (Chamorro-Premuzic, 2007) and the Basic Combat Training environment represents a complex milieu of such factors. Throughout Basic Combat Training, the military instructor cadre is working to actively shape and influence the mood state of their Soldiers as a means of achieving organizational training objectives. Many factors, such as leader-subordinate and peer-to-peer dynamics, unit morale, and individual perceptions of acute physical and mental stressors, likely contributed to differences in subjective mood among participants. Given the aggregate of observed and unobserved factors in this study, the relationship between sleep and subjective mood was most likely reduced to having a small, but still measurable, effect size. Additionally, while the phase-delayed sleep schedule resulted in increased total daily sleep for participants in the intervention group, the shortfall in daily sleep relative to known adolescent sleep needs for both groups was still large (i.e., on the order of 3–4 hours). Consequently, participants in both groups may have had a significant partial sleep deprivation that then blunted the observed effect of the schedule intervention. Finally, the phase-delayed sleep schedule, while a marked improvement over the standard Basic Combat Training sleep schedule in terms of accommodating adolescent sleep-wake patterns, was still significantly out of phase with participants' baseline patterns as inferred from participant responses on the pre-training Pittsburgh Sleep Quality Index. Such an assertion is supported by Carskadon's (2001) study of adolescent students, which found that school start times around 7 a.m. were difficult for adolescent students, and students tended to do better when start times were delayed until 8 a.m. or later.

3. Basic Rifle Marksmanship

Hypothesis 3 predicted that participants on the modified sleep schedule would exhibit greater improvement in basic rifle marksmanship scores than those following the standard Basic Combat Training sleep schedule. This hypothesis was supported by the study results, although the analysis of marksmanship performance turned out to be far from straightforward given differences between training companies in initial performance on the first record fire and variability in the number of record fires accomplished by each participant. Despite all this variability, however, it was possible to demonstrate that the degree of improvement in marksmanship performance over the serial record fires was significantly predicted, in part, by a sleep-related variable. Moreover, the effect size of sleep, while relatively small, was still greater than that attributable to prior experience with firearms.

It is noteworthy that sleep during the week preceding the record fires, when basic marksmanship tasks and subtasks were being learned, was more strongly correlated with subsequent performance than sleep during the week of the record fires. This suggests the possibility that sleep was acting as a modifier of training effectiveness. Such an assertion is consistent with research showing that procedural memories improve with subsequent early slow wave sleep (SWS) and late rapid eye movement (REM) sleep, although there is some debate regarding the relative importance of the various stages of sleep. Nevertheless, increasing evidence supports the role of sleep in memory consolidation and latent learning (Fenn, Nusbaum, & Margoliash, 2003; Gais et al., 2000; Karni et al., 1994; Stickgold, James, & Hobson, 2000; Walker et al., 2003; Wilson & McNaughton, 1994). For example, Gais and colleagues (2000) observed that memories are, on average, more than three times improved after sleep containing both SWS and REM sleep than after a period of early sleep alone. Thus, the phase-delayed schedule, which was associated with increased total daily sleep, likely increased the opportunity for late REM sleep and thereby potentiated the learning and recall of marksmanship skills.

4. Physical Fitness

Hypothesis 4 predicted that participants on the modified sleep schedule would exhibit greater improvement in physical fitness scores than participants following the standard Basic Combat Training sleep schedule. This hypothesis was not supported by the study results. As in the case of the marksmanship data, the use of nonrandomized groups led to significant baseline differences between the intervention and comparison groups, with the intervention group exhibiting higher physical fitness scores early in training. However, these differences diminished over the course of training such that the groups were equivalent on the final physical fitness assessment. Thus, the overall pattern suggested a regression to the mean phenomenon—an assertion that is supported by the absence of any correlation between fitness scores and average total daily sleep for participants in the actigraphy subsample. On the flip side, altering the timing of physical fitness training to accommodate the change in timing of sleep did not appear to harm the performance of participants in the intervention group. Additionally, participants in the intervention group generally expressed a preference for the later timing of their physical fitness training, while participants in the comparison group, on average, preferred the earlier timing of their physical fitness training.

These findings are consistent with that reported in the scientific literature examining the effect of sleep deprivation on exercise performance. Studies of exercise performance after periods of sleep deprivation of up to 72 hours have consistently demonstrated that muscle strength and exercise performance are not affected (Martin, 1981; Martin & Gaddis, 1981; Reilly & Deykin, 1983; Van Helder & Radomski, 1989). While Martin (1981) was able to show that sleep loss reduced work time to exhaustion by an average of 11 percent, this change was attributed to the psychological effects of acute sleep debt because subjects' ratings of exertion were dissociated from any cardiovascular changes. A smaller body of research has also examined the influence of chronotype on diurnal changes in muscle strength. For example, Tamm and colleagues (2009) found that evening chronotype individuals could produce a stronger maximum voluntary muscle contraction in the evening, while morning chronotype individuals exhibited no significant

change in strength throughout the day. However, the results of this study failed to show any significant effect of chronotype for the strength-based fitness assessments.

5. Sleep Hygiene

Hypothesis 5 predicted that for participants whose sleep schedules were modified, the odds of reporting occupationally significant fatigue (defined as an Epworth Sleepiness Scale (ESS) score greater than ten) would be lower than that for participants following the standard Basic Combat Training sleep schedule. This hypothesis was supported by the study results, with participants in the comparison group being 2.3 times more likely to have occupationally significant fatigue at the end of training—a finding with important safety and health implications. At the beginning of the study, participants in the intervention and comparison groups had comparable subjective sleepiness as assessed based on ESS scores. Over the course of training, participants in the comparison group exhibited a significant increase in reported sleepiness, while those in the intervention group reported no change in subjective sleepiness. Overall, evening chronotype participants reported greater sleepiness than morning chronotype participants. This result suggests that the modified sleep schedule, while an improvement over the standard schedule, still did not fully accommodate the developmental phase-delay of the adolescent and young adult circadian cycle.

Hypothesis 6 predicted that for participants whose sleep schedules were modified, the odds of reporting poor sleep quality (defined as Pittsburgh Sleep Quality Index (PSQI) score greater than five) would be lower than that for participants following the standard Basic Combat Training sleep schedule. This hypothesis was supported by the study results, with participants in the comparison group being 5.5 times more likely to report poor sleep quality at the end of training. Participants in the intervention and comparison groups had comparable sleep quality as assessed based on PSQI score at the start of the study. Over the course of training, participants in the comparison group exhibited a significant degradation in sleep quality, while those in the intervention group exhibited a trend towards improved sleep quality. Additionally, the odds of participants reporting poor quality sleep actually decreased for those in the intervention group relative

to the start of the study. This finding suggests that the phase-delayed sleep schedule was an improvement over participants' baseline sleep schedule—or in other words, Basic Combat Training actually improved the sleep hygiene of participants in the intervention group.

To summarize, participants in the intervention group graduating from Basic Combat Training did so in a better physiological state than their counterparts in the comparison group. The operational significance of this finding can be inferred from research on school age adolescents linking sleep patterns and academic performance (Acebo & Carskadon, 2001; Wolfson & Carskadon, 2003). Thus, participants in the intervention group, by way of having improved wake-sleep patterns and increased total daily sleep, were better prepared to undertake the more academically rigorous secondary military occupation-specific training that follows Basic Combat Training. Additionally, they can be expected to be at lower risk for future lost training days or injuries (Acebo, Wolfson, & Carskadon, 1997).

6. Attrition

Hypothesis 7 predicted that for participants on the modified sleep schedule, the odds of attriting from training would be lower than that for participants following the standard Basic Combat Training sleep schedule. This hypothesis was not supported by the study results as evidenced by the absence of treatment condition in the final logistic model for attrition. The single largest risk factor for attrition was sex with females more likely to attrite, followed by body mass index (i.e., fitness), neurotic personality characteristics, and depressed subjective mood. Given that the frequency of attrition relative to time was positively skewed—that is, most attrition tends to occur earlier rather than later in training—it is more likely that pre-existing conditions or vulnerabilities were the predominant determinant of attrition.

F. SELECT HUMAN SYSTEMS INTEGRATION ANALYSES

Up to this point, we have described a research study that was conducted from the behavioral sciences paradigm utilizing an experimental methodology and multi-variable statistical techniques drawn from experimental psychology. We proposed a series of

research hypotheses and developed corresponding statistical models to aid decision making with regards to our accepting or rejecting those research hypotheses (and conversely their corollary null hypotheses). However, if we are to transition from the behavioral sciences to the HSI paradigm, we need to take a subset of research hypotheses that were accepted based on the statistical models and reformulate those that are of most interest to us in terms of tradeoff functions, thereby making possible their direct incorporation in the "system analytic thinking process" (Weisz, 1967, p. 3). The latter is involved whenever there is a choice between various alternative system mixes to meet a particular requirement or threat. Historically, systems analysis has been dominated by mathematically based operations research techniques developed to facilitate the decision making of organizational planners and systems developers (Hughes, 1998). Consequently, the objective of our forthcoming HSI analyses is the development of mathematical tradeoff functions that can then be used by decision makers to predict the optimum mix of human performance determinants, whether in terms of cost, effectiveness, or technical feasibility (Weisz, 1967, 1968). This objective will be accomplished using the isoperformance methodology (Jones & Kennedy, 1996) described in depth in Chapter IV. In so doing, we establish the pattern by which human factors research and human considerations can be appropriately represented in systems analyses.

1. Basic Rifle Marksmanship Model

The purpose of this section is to develop in a step-by-step fashion an isoperformance curve for basic rifle marksmanship. We start with a model, a criterion level, and a confidence level. The model states the functional dependence of marksmanship performance on aptitude and average daily sleep. The criterion indicates the minimal level of performance that one is willing to regard as adequate. The confidence level is the probability of adequate performance, by which we mean that performance will equal or exceed the criterion. What results is essentially a tradeoff function for marksmanship in terms of the personnel (i.e., aptitude) and survivability (i.e., fatigue) domains of HSI.

Our first step is to obtain an expression for a model for the expected marksmanship performance for an individual Soldier, *i*. As will be recalled from our earlier analysis of the basic rifle marksmanship data for the actigraphy subsample, participants in the intervention group tended to have lower initial marksmanship scores relative to participants in the comparison group, but they also exhibited a greater improvement in marksmanship performance over serial firings. Additionally, the magnitude of this change was positively correlated with average daily sleep during the week prior to the serial firings ($\rho = 0.341$, $p = 0.001$), which was when they received instruction in rifle marksmanship fundamentals. Moreover, there was no effect of group when sleep was included in the analysis, implying that differences in instructor cadre were not a likely explanation for the observed difference in basic rifle marksmanship. Consequently, we propose the following model for the basic rifle marksmanship data:

$$\Delta \mathrm{S}_i = a + b\left(\mathrm{SLP}_i\right) + \varepsilon_i \tag{5}$$

where $\Delta \mathrm{S}_i$ is the difference between first and last serial marksmanship scores for the i^{th} Soldier, and SLP_i is the i^{th} Soldier's average daily sleep during the week prior to the serial firings. The constants, *a* and *b*, are parameters estimated during the model fitting and ε_i is a normally distributed error term with mean equal to zero and variance equal to σ_ε^2.

Table VI-39 presents a conventional readout for the model in terms of expected mean squares, F ratio, significance level, and effect size. The result is that average daily sleep is a significant determinant of $\Delta \mathrm{S}$, explaining nearly 11% of the variance in the change in marksmanship scores. While average daily sleep has a relatively modest effect on marksmanship performance, it is a determinant that is, at least in Basic Combat Training, controllable by the Army.

Table VI-39. Expected mean squares, F ratio, significance level, and effect size for the basic rifle marksmanship data from the actigraphy subsample.

Source	MS	df	F	p	η^2
Sleep	412.190	1	11.329	0.001	0.116
Error	36.382	86	—	—	—

We can rewrite Equation 5 as follows:

$$E\left[\Delta \mathrm{S}_i\right] = a + b\left(\mathrm{SLP}_i\right) \tag{6}$$

The only difference between the right side of this equation and that of the full model is the absence of the error term. Hence, the expected change in marksmanship performance for the i^{th} Soldier depends only on the determinant SLP_i. The next step is to modify the model so that the left hand side of Equation 6 is in terms of the excepted final marksmanship score. We begin by noting that $\Delta \mathrm{S}_i = \mathrm{S}_{2i} - \mathrm{S}_{1i}$, where S_{1i} is a Soldier's initial marksmanship score and S_{2i} is their final marksmanship score. According, we rewrite Equation 6:

$$E\left[\mathrm{S}_{2i} - \mathrm{S}_{1i}\right] = a + b\left(\mathrm{SLP}_i\right) \tag{7}$$

Since expectation of a difference is simply the difference of expectations:

$$E\left[\mathrm{S}_{2i}\right] - E\left[\mathrm{S}_{1i}\right] = a + b\left(\mathrm{SLP}_i\right) \tag{8}$$

Rearranging terms:

$$E\left[\mathrm{S}_{2i}\right] = E\left[\mathrm{S}_{1i}\right] + a + b\left(\mathrm{SLP}_i\right) \tag{9}$$

We next propose replacing the $E\left[\mathrm{S}_{1i}\right]$ term with $E\left[\mathrm{S}_{1j}\right]$, which is the expectation of the initial marksmanship score for a Soldier in the j^{th} quintile for initial marksmanship performance. Consequently, Equation 9 becomes

$$E\left[\mathrm{S}_{2ij}\right] = E\left[\mathrm{S}_{1j}\right] + a + b\left(\mathrm{SLP}_{ij}\right) \tag{10}$$

which requires that we recalculate σ_ε^2. It is observed that the penalty for this change is small, with σ_ε^2 now equal to 37.049 as compared to 36.382 previously.

We explain further since it may not be intuitive why we have proceeded through the following model development steps rather than simply fitting a model directly using S_{1ij}, SLP_{ij}, and S_{2ij}. To start, it was observed that there was a strong multi-collinearity between S_{1ij} and SLP_{ij}, which complicates attempts at regression analysis. Another nontrivial problem encountered in this study was the finding that the intervention and comparison groups differed in terms of initial marksmanship performance, and hence, aptitude—an observation that can be attributed to the use of non-randomly formed groups in the study design. Since the intervention group, which obtained more sleep by study design, had worse initial marksmanship performance, sleep is negatively correlated with initial marksmanship scores (i.e., the effect of sleep was confounded by group differences in aptitude). However, as we showed earlier in this section, sleep is also positively correlated with improvement in serial marksmanship scores irrespective of group. These are contradictory findings. If sleep did indeed have a negative effect on initial marksmanship performance, it would be expected to have a negative effect on serial marksmanship performance as well—but exactly the opposite was observed. Thus, we focused on fitting the latter relationship to minimize potential confounding by the former. In the end, however, we still need to express the model dependent variable in terms of final marksmanship scores as this is the performance criterion used by the Army.

The second step in developing the isoperformance curve is to determine what expected performance for the i^{th} Soldier in the j^{th} quintile must be if the probability of adequate performance is to equal a specified confidence interval. In our case, the Army has specified a final marksmanship score of 23 as the criterion, and we will presuppose 0.80 is the desired confidence level. These specifications are met if the expected performance for the i^{th} Soldier in the j^{th} quintile is

$$E\left[\mathrm{S}_{2ij}\right] = 23 + z\sigma_{\varepsilon} \tag{11}$$

where z equals 0.84 from tables of the normal curve and $\sigma_\varepsilon = \sqrt{37.049}$ (see prior paragraphs).[22] Hence,

$$E\left[\mathrm{S}_{2ij}\right] = 23 + 0.84\sqrt{37.049} = 28.11 \tag{12}$$

If the final marksmanship score for the i^{th} Soldier in the j^{th} quintile is to equal or exceed 23 with a probability of 0.80, then the expected final marksmanship score for the Soldier must equal 28.11.

The third and last step is to put Equations 10 and 12 together. Doing so produces

$$28.11 = E\left[\mathrm{S}_{1j}\right] + a + b\left(\mathrm{SLP}_{ij}\right) \tag{13}$$

Equation 13 involves two model parameters (a and b), five sample statistics $\left(E\left[\mathrm{S}_{1j}\right]\right)$, and the determinants SLP_i and quintile j, the latter corresponding to a choice of aptitude level. Rearranging terms so that SLP_{ij} is on the left hand side, one obtains

$$\mathrm{SLP}_{ij} = \frac{28.11 - E\left[\mathrm{S}_{1j}\right] - a}{b} \tag{14}$$

The estimated values for the model parameters and sample statistics in Equation 14 are given below:

$$\begin{array}{ll} a = -19.052 & E\left[S_{1,1}\right] = 12.250 \\ b = 3.861 & E\left[S_{1,2}\right] = 18.000 \\ & E\left[S_{1,3}\right] = 23.579 \\ & E\left[S_{1,4}\right] = 27.056 \\ & E\left[S_{1,5}\right] = 31.053 \end{array}$$

This is the basic rifle marksmanship isoperformance curve. For any given choice of aptitude quintile, j, one can now calculate a value of SLP_{ij} such that the two together produce adequate performance with the specified level of confidence.

[22] For the sake of simplicity of illustration, we fit a confidence interval using the procedure described by Jones and Kennedy (1996). As was discussed in Chapter V, a more conservative trade off analysis would be obtained by instead fitting the prediction interval, which accounts for the uncertainty present in the estimates of the model parameters.

Figure VI-54 presents three isoperformance curves that trade off aptitude, as assessed based on initial marksmanship score, and average daily sleep. The criterion is set at 23 (i.e., the minimum marksmanship qualification threshold), 27, and 30 (i.e., the sharp shooter qualification threshold). Each isoperformance curve traces combinations of aptitude and average daily sleep that yield equivalent performance in terms of the criterion, which in this case is final marksmanship score. Thus, these isoperformance curves can be read as tradeoff functions. For example, Soldiers sleeping 7.55 hours per day will meet the basic rifle marksmanship qualification threshold of a final score of 23 if their initial marksmanship score is at least 18. Alternatively, if Soldiers are allowed to sleep for only 6.77 hours per day, then their initial marksmanship score will need to be at least 21 if they are to achieve the basic rifle marksmanship criterion on their final record fire. In other words, it takes one point in marksmanship aptitude to make up for each 16 minute reduction in Soldiers' average daily sleep during marksmanship instruction.

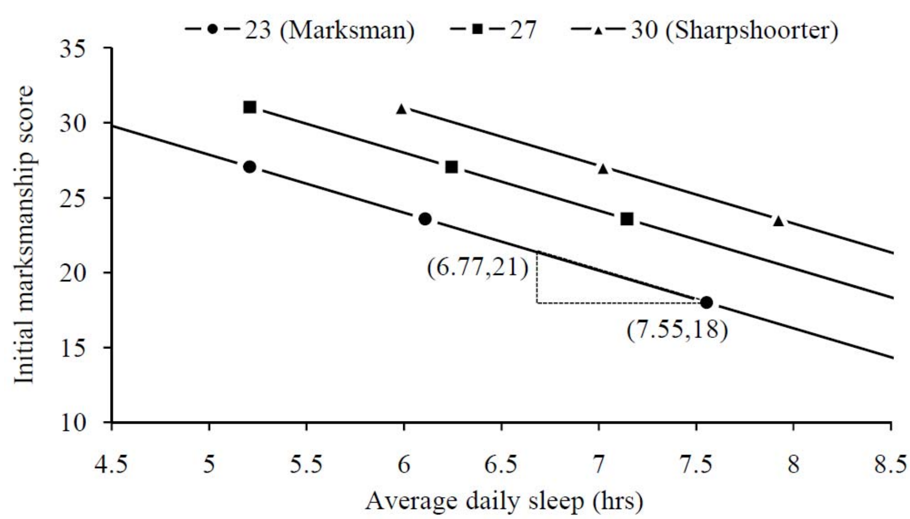


Figure VI-54. Isoperformance curves trading off aptitude, expressed as initial marksmanship score, and average daily sleep, setting the final marksmanship score criterion levels at 23, 27, and 30 and percentage proficient at 80%.

2. Sleep Quality Model

Repeating the process used to create the marksmanship isoperformance model, we next develop an isoperformance curve for post-training sleep quality as assessed using the Pittsburgh Sleep Quality Index (PSQI). Again, we start with a model, a criterion level, and a confidence level. The model states the functional dependence of post-training sleep quality on pre-training sleep quality and average daily sleep during training. Since sleep quality is an important clinical construct and poor sleep quality is a significant symptom of many medical, psychiatric, and sleep disorders (Buysse et al., 1988), we utilize PSQI scores as a metric for the occupational health domain of HSI. In terms of a criterion, a global PSQI score of greater than 5 was shown by Buysse and colleagues (1988) to have a 90% diagnostic sensitivity in distinguishing good sleepers (i.e., healthy individuals) from poor sleepers (i.e., individuals with mood or sleep disorders). What results is essentially a tradeoff function in terms of the personnel (i.e., individuals' baseline sleep quality) and survivability (i.e., fatigue) domains of HSI.

Our first step is to obtain an expression for a model of the expected post-training sleep quality of an individual Soldier, *i*. We propose the following model for the post-training PSQI data:

$$\mathrm{PSQI}_{2i} = a + b\left(\mathrm{PSQI}_{1i}\right) + c\left(\mathrm{SLP}_i\right) + \varepsilon_i \qquad (15)$$

where PSQI_{2i} is the post-training PSQI score for the i^{th} Soldier, PSQI_{1i} is the i^{th} Soldier's baseline PSQI score prior to starting training, and SLP_i is the i^{th} Soldier's average daily sleep during training. The constants, *a*, *b*, and *c*, are parameters estimated during the model fitting and ε_i is a normally distributed error term with mean equal to zero and variance equal to σ_ε^2. Table VI-40 presents a conventional readout for the model in terms of expected mean squares, F ratio, significance level, and effect size. The result is that both baseline sleep quality and average daily sleep are significant determinants of PSQI_2.

Table VI-40. Expected mean squares, F ratio, significance level, and effect size for the post-training PSQI score data from the actigraphy subsample.

Source	MS	df	*F*	*p*	η^2
Baseline PSQI score	39.152	1	4.264	0.043	0.057
Sleep	64.233	1	6.995	0.010	0.090
Error	9.183	71	—	—	—

We can rewrite Equation 15 as follows:

$$E\left[\mathrm{PSQI}_{2i}\right] = a + b\left(\mathrm{PSQI}_{1i}\right) + c\left(\mathrm{SLP}_i\right) \tag{16}$$

The only difference between the right side of this equation and that of the full model is the absence of the error term. Hence, the expected post-training PSQI score for the i^{th} Soldier depends only of the determinants PSQI_{1i} and SLP_i.

The second step in developing the isoperformance curve is to determine what the expected post-training PSQI score for the i^{th} Soldier must be if the probability of adequate sleep quality is to equal a specified confidence interval. In this case, we use the cutoff global PSQI score of 5 suggested by Buysse and colleagues (1988) as the criterion, and we will presuppose 0.80 is the desired confidence level. These specifications are met if the expected post-training PSQI score for the i^{th} Soldier is

$$E\left[\mathrm{PSQI}_{2i}\right] = 5 - z_{0.80}\sigma_{\varepsilon} \tag{17}$$

where z equals 0.84 from tables of the normal curve and $\sigma_{\varepsilon} = \sqrt{9.183}$. Hence,

$$E\left[\mathrm{PSQI}_{2i}\right] = 5 - 0.84\sqrt{9.183} = 2.455 \tag{18}$$

If the post-training PSQI score for the i^{th} Soldier is less than or equal to 5 with a probability of 0.80, then the expected post-training PSQI score for the Soldier must equal 2.455.

The third and last step is to put Equations 16 and 18 together. Doing so produces

$$2.455 = a + b\left(\mathrm{PSQI}_{1i}\right) + c\left(\mathrm{SLP}_i\right) \tag{19}$$

Rearranging terms so that SLP_i is on the left hand side, one obtains

$$\mathrm{SLP}_i = \frac{2.455 - a - b\left(\mathrm{PSQI}_{1i}\right)}{c} \tag{20}$$

The estimated values for the model parameters in Equation 20 are given below:

$$a = 16.129$$
$$b = 0.296$$
$$c = -2.053$$

This is the post-training sleep quality isoperformance curve. For any given choice of baseline sleep quality, PSQI_{1i}, one can now calculate a value of SLP_i such that the two together produce adequate post-training sleep quality (i.e., occupational health) with the specified level of confidence.

Figure VI-55 presents two isoperformance curves that trade off baseline sleep quality and average daily sleep during training. The criterion is set at 5, the clinical threshold for healthy individuals, and 6.5, the average baseline PSQI score in the study sample. The latter criterion setting represents the option of "doing no harm"—that is, not further exacerbating the sleep quality of already poor sleepers. Each isoperformance curve traces combinations of baseline sleep quality and average daily sleep that yield equivalent performance in terms of the criterion, which in this case is post-training sleep quality. Consequently, these isoperformance curves can be read as tradeoff functions. For example, Soldiers with poor baseline sleep quality (e.g., PSQI = 9.1) can obtain good sleep quality if they are provided 7.98 hours of sleep per night during training. Alternatively, if Soldiers are allowed to sleep for only 7.22 hours per day, then their baseline sleep quality will need to be fairly good (e.g., PSQI = 3.9) if they are to achieve the post-training sleep quality criterion. In other words, it takes one point in baseline PSQI score to make up for each 9 minutes reduction in Soldiers' average daily sleep during training.

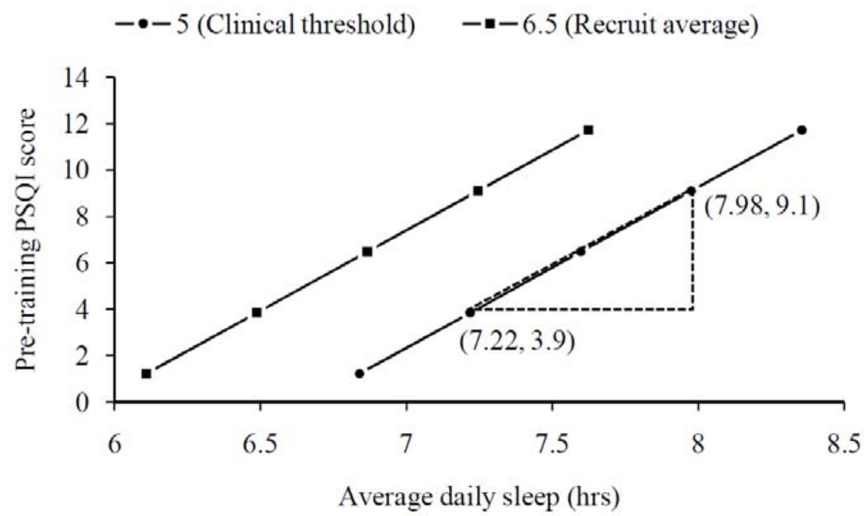


Figure VI-55. Isoperformance curves trading off baseline sleep quality, expressed as pre-training PSQI score, and average daily sleep, setting the final PSQI score criterion levels at 5 and 6.5 and the assurance level at 80%.

G. CONCLUSION

In summary, increasing sleep and concomitantly decreasing fatigue had a small but measurable influence on various indicators of Soldier functioning even after controlling for a variety of factors that affect performance. Although Soldiers' responses to the phase-delayed schedule intervention were relatively modest, it should be appreciated that the majority of outcome measures in Basic Combat Training are not highly sensitive to the effects of fatigue. Thus, the most important finding of the study may be the impact of the schedule intervention on sleep quality during Basic Combat Training—that is, Soldiers completing Basic Combat Training using the phase-delayed sleep schedule had significant improvements in sleep hygiene such that they graduated from training in a better physiological state than when they started. Or, in other words, the phase-delayed sleep schedule allowed Soldiers to accomplish the training objectives of Basic Combat Training at a lower cost in terms of their sleep reservoir, thereby leaving them with a greater available cognitive work capacity going forward for subsequent training. The significance of this finding may not be fully appreciated until Soldiers'

subsequent performance is assessed during the more cognitively demanding secondary military occupational specialty training courses—a recommendation for follow-up research related to this work.

While insufficient sleep and the consequent fatigue is a recognized problem in our society, concern has mainly been voiced around well publicized, high cost disasters resulting from the degraded occupational performance of sleep-deprived adults. The role of sleep in less dramatic circumstances seems to be underappreciated, particularly in the military environment where inadequate sleep is considered part and parcel of the routine starting in basic military training and onward. To the extent that adolescents and young adults entering the Army are unable to obtain sufficient sleep at the appropriate time to facilitate their primary developmental task—that being to master core Soldiering skills and incorporate Army values within their evolving self-identity—there are potentially significant hidden lost opportunity costs being borne by the Army. Our HSI tradeoff analyses, derived from the results of a behavioral sciences experiment involving a simple sleep schedule intervention, provide an empirical foundation to begin quantitatively assessing the contribution of sleep to Soldier well-being and performance. What should then emerge is a *Weltanschauung* that considers the human sleep reservoir in terms of its contribution to the performance of the human component of weapon systems or the human as a weapon system. Accordingly, the quantity and quality of sleep become limited resource variables that can and must be considered as part of the human factors contribution to systems analyses.

H. BIBLIOGRAPHY

Acebo, C., & Carskadon, M. A. (2001). Influence of irregular sleep patterns on waking behavior. In M. A. Carskadon (Ed.), *Adolescent sleep patterns: biological, social, and psychological influences* (chap 13, pp. 220–235). New York, NY: Cambridge University Press.

Acebo, C., Wolfson, A. R., & Carskadon, M. A. (1997). Relations among self-reported sleep patterns, health, and injuries in adolescents. *Sleep Research*, 26, 149.

Anch, A. M., Browman, C. P., Mitler, M., & Walsh, J. K. (1988). *Sleep: a scientific perspective*. Englewood Cliffs, NJ: Pearson Prentice-Hall, Inc.

Andrews, C. H. (2004). *The relationship between sleep regimen and performance in United States Navy recruits.* Monterey, CA: Naval Postgraduate School.

Baldus, B. R. (2002). *Sleep patterns in U.S. Navy recruits: an assessment of the impact of changing sleep regimens*. Monterey, CA: Naval Postgraduate School.

Belenky, G., Wesensten, N. J., Thorne, D. R., Thomas, M. L., Sing, H. C., Redmond, D. P., et al. (2003). Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. *Journal of Sleep Research*, 12, 1–12.

Birchler-Pedross, A., Schröder, C. M., Münch, M., Knoblauch, V., Blatter, K., Schnitzler-Sack, C., Wirz-Justice, A., & Cajochen, C. (2009). Subjective well-being is modulated by circadian phase, sleep pressure, age, and gender. *Journal of Biological Rhythms*, 24(3), 232–242.

Boivin, D. B., Czeisler, C. A., Dijk, D. J., Duffy, J. F., Folkard, S., Minors, D. S., Totterdell, P., & Waterhouse, J. M. (1997). Complex interaction of the sleep-wake cycle and circadian phase modulates mood in healthy subjects. *Archives of General Psychiatry*, 54, 145–152

Booher, H. R. (2003). Introduction: human systems integration. In H. R. Booher (Ed.), *Handbook of human systems integration* (chap. 1, pp. 1–30). New York, NY: John Wiley and Sons.

Brick, J. M., & Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research*, 5, 215–238.

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Research*, 28(2), 193–213.

Carskadon, M. A. (2001). Factors influencing sleep patterns of adolescents. In M. A. Carskadon (Ed.), *Adolescent sleep patterns: biological, social, and psychological influences* (chap 2, pp. 4–26). New York, NY: Cambridge University Press.

Carskadon, M. A., Acebo, C., Richardson, G. S., Tate, B. A., & Seifer, R. (1997). An approach to studying circadian rhythms of adolescent humans. *Journal of Biological Rhythms*, 12, 278−289.

Carskadon, M. A., Wolfson, A. R., Acebo, C., Tzischinsky, O., & Seifer, R. (1998). Adolescent sleep patterns, circadian timing, and sleepiness at a transition to early school days. *Sleep*, 21(8), 871–881.

Chamorro-Premuzic, T. (2007). *Personality and individual differences*. Malden, MA: Blackwell Publishing.

Colton, C. C. (1837). *Lacon, or many things in few words*. London, UK: Longman, Orme, Brown, Green, and Longmans.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO personality inventory (NEO-PI-R) and NEO five-factor inventory (NEO-FFI) professional manual.* Odessa, FL: Psychological Assessment Resources.

Danilenko, K. V., Cajochen, C., & Wirz-Justice, A. (2003). Is sleep per se a zeitgeber in humans? *Journal of Biological Rhythms*, 18, 170–178.

Deal, S. (2007, January). *Definition of human systems integration*. HSI working group meeting conducted at International Council of Systems Engineering International Workshop 2007, Albuquerque, New Mexico.

Defense Acquisition University (2009). Human systems integration. *Interim defense acquisition guidebook* (chap. 6). Ft. Belvoir, VA: Defense Acquisition University Press.

Department of Defense (2008). *Department of Defense instruction 5000.02: operation of the defense acquisition system*. Washington, DC: Author.

Department of Defense (2009). *Chairman of the Joint Chiefs of Staff instruction 3170.01G: joint capabilities integration and development system*. Washington, DC: Author.

Deuster, P. A., O'Connor, F. G., Henry, K. A., Martindale, V. E., Talbot, L., Jonas, W., & Friedl, K. (2007). Human performance optimization: an evolving charge to the department of defense. *Military Medicine*, 172(11), 1133–1137.

Directorate Basic Combat Training Doctrine and Training Development (2008, Mar). *Program of instruction for basic combat training* (Version 5). Fort Jackson, SC: Author.

Driskell, J.E., Hughes, S.C., Willis, R.C., Cannon-Bowers, J., & Salas. E. (1991). *Stress, stressor, and decision-making* (Technical Report for the Naval Training Systems Center). Orlando, FL: Naval Training Systems Center.

Driskell, J.E., & Salas, E. (Eds.) (1996). *Stress and human performance*. Mahwah, NJ: Lawrence Erlbaum Associates.

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature*, 425, 614–616.

Gais, S., Plihal, W., Wagner, U., & Born, J. (2000). Early sleep triggers memory for early visual discrimination skills. *Nature Neuroscience*, 3(12), 1335–1339.

Graham, M. G. (Ed.), National Research Council and Institute of Medicine (2000). *Sleep needs, patterns, and difficulties of adolescents.* Washington, DC: National Academy Press.

Horne, J. A. & Ostberg, O. (1976). A self assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology*, 4, 97–110.

Hughes, T. P. (1998). *Rescuing Prometheus*. New York, NY: Vintage Books.

Hursh, S.R., & Bell, G.B. (2001). *Human performance cognitive model for Air Force information operations* (Technical Report for the Air Force Research Laboratory). Brooks Air Force Base, TX: Air Force Research Laboratory.

Hursh, S. R., Redmond, D. P., Johnson, M. L., Thorne, D. R., Belenky, G., Balkin, T. J., Storm, W. F., Miller, J. C., & Eddy, D. R. (2004). Fatigue models for applied research in warfighting. *Aviation, Space, and Environmental Medicine*, 75(3 Suppl), A44–A53.

Johns, M. W. (1991). A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep*, 14, 540–545.

Johnson, D. C., Polusny, M. A., Erbes, C., King, D., King, L., Litz, B., Schnurr, P., Friedman, M., & Southwick, S. (2008). *The response to stressful experiences scale (RSES)*. U.S. Department of Veterans Affairs National Center for PTSD.

Jones, M. B., & Kennedy, R. S. (1996). Isoperformance curves in applied physiology. *Human Factors*, 38(1), 167–182.

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J. M., & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of perceptual skill. *Science*, 265, 679–682.

Killgore, W. D., Estrada, A., Wildzunas, R. M., & Balkin, T. J. (2008). *Sleep and performance measures in soldiers undergoing military relevant training.* Presented at the 26th Army Science Conference, Orlando, FL.

Killgore, W. D., Balkin, T. J., & Wesensten, N. J. (2006). Impaired decision making following 49 h of sleep deprivation. *Journal of Sleep Research, 15(1)*, 7–13.

Kim, J. O., & Curry, J. (1977). The treatment of missing data in multivariate analysis. *Sociological Methods and Research*, 6, 216–240

Krueger, G. P. (1990). Sustained military performance in continuous operations: combatant fatigue, rest, and sleep needs. In R. Gal & A. D. Mangelsdorff (Eds.), *Handbook of military psychology* (pp. 244–277). Chichester, UK: John Wiley and Sons.

Lange, T., Perras, B., Fehm, H. L., & Born, J. (2003). Sleep enhances the human antibody response to Hepatitis A vaccination. *Psychosomatic Medicine*, 65, 831–835.

Martin, B. J. (1981). Effect of sleep deprivation on tolerance of prolonged exercise. *European Journal of Applied Physiology*, 47, 345–354.

Martin, B. J., & Gaddis, G. M. (1981). Exercise after sleep deprivation. *Medicine and Science in Sports and Exercise*, 13(4), 220–223.

Matthews, G., Davies, D. R., Westerman, S. J., & Stammers, R. B. (2000). *Human performance: cognition, stress, and individual differences*. Philadelphia, PA: Taylor and Francis.

McLellan, T. M., Kamimori, G. H., Bell, D. G., Smith, I. F., Johnson, D., and Belenky, G. (2005). Caffeine maintains vigilance and marksmanship in simulated urban operations with sleep deprivation. *Aviation, Space, and Environmental Medicine*, 76(1), 39–45.

McNair, D. M., Lorr, M., & Droppleman, L. F. (1981). *Manual for the profile of mood states.* San Diego, CA: Educational and Industrial Testing Services.

Mercer, P. W., Merritt, S. L., & Cowell, J. M. (1998). Differences in reported sleep need among adolescents. *Journal of Adolescent Health,* 23(5), 259–263.

Miller, D. B. (2005). *Sleep and predicted cognitive performance of new cadets during cadet basic training at the United States Military Academy.* Monterey, CA: Naval Postgraduate School.

Miller, J. C. (2006). *Fundamentals of shiftwork scheduling* (Report No. AFRL-HE-BR-2006-0011). Brooks City-Base, TX: Human Effectiveness Directorate, Air Force Research Laboratory.

Miller, N. L., Matsangas, P., & Shattuck, L. G. (2007). Fatigue and its effect on performance in military environments. In P. A. Hancock & J. L. Szalma (Eds.), *Performance under stress* (chap. 12, pp. 231–249). Burlington, VT: Ashgate Publishing Company.

Miller, N. L., & Shattuck, L. G. (2005). Sleep patterns of young men and women enrolled at the United States Military Academy: Results from year one of a four year longitudinal study. *Sleep*, 28(7), 837–841.

Miller, N. L., Shattuck, L. G., Panagiotis, M., & Dyche, J. (2008). Sleep and academic performance in U.S. military training and education programs. *Mind, Brain, and Education*, 2(1), 29–33.

Moldofsky, H. (1995). Sleep and the immune system. *International Journal of Immunopharmacology*, 17(8), 649–654.

Monk, T. H., Buysse, D. J., Reynolds, C. F., Jarrett, D. B., & Kupfer, D. J. (1992). Rhythmic vs. homeostatic influences on mood, activation, and performance in young and old men. *Journal of Gerontology*, 47(4), 221–227.

Monk, T. H., Buysse, D. J., & Rose, L. R. (1999). Wrist actigraphic measures of sleep in space. *Sleep*, 22(7), 948–954.

Naval Center for Combat and Operational Stress Control (2009). *Resilience: what is it?* Retrieved July 16, 2009, from http://www.med.navy.mil/sites/nmcsd/nccosc/Resilience/Documents/Resilience_What_Is_It.pdf

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (chap. 1, pp. 1–53). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Paquet, J., Kawinska, A., & Carrier, J. (2007). Wake detection capacity of actigraphy during sleep. *Sleep*, 30(10), 1362–1369.

Pew, R. W., & Mavor, A. S. (Eds.) (2007). *Human-system integration in the system development process: a new look.* Washington, DC: National Academies Press.

Pigeau, R., Angus, B., & O'Neil, P. (1995). Vigilance latencies to aircraft detection among NORAD surveillance operators. *Human Factors*, 37(3), 622–634.

Reilly, T., & Deykin, T. (1983). Effects of partial sleep loss on subjective states, psychomotor and physical performance tests. *Journal of Human Movement Studies*, 9, 157–170.

Rosa, R. H. (2001). What can the study of work scheduling tell us about adolescent sleep? In M. A. Carskadon (Ed.), *Adolescent sleep patterns: biological, social, and psychological influences* (chap 10, pp. 159–171). New York, NY: Cambridge University Press.

Selvi, Y., Gulec, M., Agargun, M. Y., & Besiroglu, L. (2007). Mood changes after sleep deprivation in morningness-eveningness chronotypes in healthy individuals. *Journal of Sleep Research*, 16, 241–244.

Snow, R. E. (1978). Theory and method for research on aptitude processes. *Intelligence, 2*, 225–278.

Stickgold, R., James, L., & Hobson, A. (2000). Visual discrimination learning requires sleep after training. *Nature Neuroscience*, 3(12), 1237–1238.

Tamm, A. S., Lagerquist, O., Ley, A. L., & Collins, D. F. (2009). Chronotype influences diurnal variations in the excitability of the human motor cortex and the ability to generate torque during a maximum voluntary contraction. *Journal of Biological Rhythms*, 24(3), 211–224.

Taub, J. M., & Berger, R. J. (1973). Performance and mood following variations in the length and timing of sleep. *Psychophysiology*, 10(6), 559–570.

Tharion, W. J., Shukitt-Hale, B., & Lieberman, H. R. (2003). Caffeine effects on marksmanship during high-stress military training with 72 hour sleep deprivation. *Aviation, Space, and Environmental Medicine*, 74(4), 309–314.

Thorne, D. R., Thomas, M. L., Russo, M.B., Sing, H. C., Balkin T. J., Wesensten, N. J., et al. (1999) Performance on a driving-simulator divided attention task during one week of restricted nightly sleep. *Sleep*, 22(Suppl 1), 301.

Trickel, M. T., Barnes, M. D., & Egget, D. L. (2000). Health-related variables and academic performance among first-year college students: implications for sleep and other behaviors. *Journal of American College Health*, 49(3), 125–131.

Tvaryanas, A. P., Brown, L., & Miller, N. L. (2009). Managing the human weapon system: a vision for an Air Force human performance doctrine. *Air and Space Power Journal*, 23(2), 34–41.

Van Dongen, H. P. A., Maislin, G., Mullington, J. M., and Dinges, D. F. (2003). The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep*, 26, 2.

Van Helder, T., & Radomski, M. W. (1989). Sleep deprivation and the effect of exercise performance. *Sports Medicine*, 7, 235–247.

Walker, M. P., Brakefield, T., Hobson, J. A., Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature*, 425, 616–620.

Weiner, I. B., & Greene, R. L. (2008). *Handbook of personality assessment*. Hoboken, NJ: John Wiley & Sons.

Weisz, J. D. (1967). *Systems analysis: human factors research and development contributions* (Technical Note 6-67). Aberdeen Proving Ground, MD: Human Engineering Laboratory. (DTIC Accession No. AD818566)

Weisz, J. D. (1968). *Systems analysis: manpower resource/system design integration methodology* (Technical Note 9-68). Aberdeen Proving Ground, MD: Human Engineering Laboratory. (DTIC Accession No. AD675481)

Whitener, E. M. (1989). A meta-analytic review of the effect on learning of the interaction between prior achievement and instructional support. *Review of Educational Research*, 59(1), 65–86.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, 265, 676–679.

Wolfson, A. R. (2001). Bridging the gap between research and practice: what will adolescents' sleep-wake patterns look like in the 21st century? In M. A. Carskadon (Ed.), *Adolescent sleep patterns: biological, social, and psychological influences* (chap 12, pp. 198–219). New York, NY: Cambridge University Press.

Wolfson, A. R., & Carskadon, M. A. (2003). Understanding adolescents' sleep patterns and school performance: a critical appraisal. *Sleep Medicine Reviews*, 7(6), 491–506.

Wood, C., & Magnello, M. E. (1992). Diurnal changes in perceptions of energy and mood. *Journal of the Royal Society of Medicine*, 85, 191–194.

Zigler, R. N, & Weiss, R. A. (2003). Personnel survivability methodology. In H. R. Booher (Ed.), *Handbook of human systems integration* (chap. 16, pp. 595–630). New York, NY: John Wiley and Sons.

THIS PAGE INTENTIONALLY LEFT BLANK

VII. HUMAN SYSTEMS INTEGRATION DOMAIN TRADEOFFS IN OPTIMIZED MANNING – THE TASK EFFECTIVENESS SCHEDULING TOOL

> Pretending to be superhuman is very dangerous. In a well-led military, the self-maintenance of the commander, the interests of his or her country, and the good of the troops are incommensurable only when the enemy succeeds in making them so. It is time to critically reexamine our love affair with stoic self-denial…If an adversary can turn our commanders into sleepwalking zombies, from a moral point of view the adversary has done nothing fundamentally different than destroying supplies of food, water, or ammunition. Such could be the outcome, despite our best efforts to counter it. But we must stop doing it to ourselves and handing the enemy a dangerous and unearned advantage (Shay, 1998, p. 104).

A. INTRODUCTION

The first mathematical models of sleep and circadian processes were developed more than 20 years ago in an effort to explain the timing of the human sleep-wake activity cycle. In the intervening years, a number of applied biomathematical models of fatigue and performance have been developed from the first generation of models of sleep-wake cycles. These applied biomathematical models typically use information about sleep history, duration of wakefulness, and circadian phase to predict performance capability and risk. They are currently used to assess the potential contribution of fatigue to performance degradation at specific points in time, to develop and evaluate work/rest schedules, to plan work and sleep in operational missions, and to determine the timing of fatigue countermeasures to anticipated performance decrements (Neri, 2004). The March 2004 edition of the journal, *Aviation, Space, and Environmental Medicine*, provides a comprehensive review and model-to-data comparisons of seven of the current biomathematical models of human fatigue and performance. Those interested in more information on the biomathematical modeling of fatigue and performance should reference this resource and the bibliographies contained within.

The U.S. Defense Department has long pursued applied research concerning fatigue in military operations and has developed several biomathematical fatigue models. One of these models, known as the Sleep, Activity, Fatigue, and Task Effectiveness

(SAFTE) Model, has achieved relatively wide acceptance and seen practical application within the Fatigue Avoidance Scheduling Tool (FAST) (Hursh et al., 2004). FAST is used by various military occupational communities in conjunction with rule-based heuristics (e.g., shift-work guidelines, hours-of-service rules, etc.) to develop plans for staffing system functions or missions. FAST is also beginning to be used by the system development community, again as an augmentation of other heuristics, to develop and refine manpower estimates in light of predictions of human performance. For instance, organizational planners may use rule-based heuristics to determine staffing needs, while ignoring potential constraints, and then iteratively refine the solution, using heuristics and FAST, to then attempt to meet constraints and satisfy objectives. The result is necessarily a trial-and-error approach that attempts to take manpower and performance into account, but does not systematically minimize manpower or maximize performance.

Such instances beg the question: do current, commercially available implementations of biomathematical models of fatigue, with FAST being an archetype, answer the questions being asked by organizational planners? In essence, the current instantiation of FAST requires the user to provide a schedule for which the software computes predicted task effectiveness over some time period of interest. Thus, given a schedule, one can get a forecast for future task effectiveness. But what about the inverse question: given a desired threshold or lower limit for task effectiveness, what is the *optimal* schedule in terms of the timing of sleep-wake periods and the assignment of performance-sensitive duties? And by extension, there is the corollary question, how many people are needed to achieve sustained performance above the desired threshold? The operational relevance of these questions should be self-evident given the current emphasis on minimal manning paradigms for many military weapon systems.

In current vernacular, FAST is a point solution because it is tailored to provide a forecast of task effectiveness for a particular schedule. As such, it cannot directly answer the aforementioned questions that are most germane to organizational planners—that is, it does not allow for a systematic exploration of a solution space to determine an optimal solution in terms of manning, schedule, or both. Consequently, the question taken up

here is the feasibility of reconciling this problem within the self-imposed constraint of using the existing implementation of the SAFTE model in FAST.

B. PROBLEM STATEMENT

To illustrate an approach to solving this problem, consider the general dynamic system represented by the block diagram in Figure VII-1. The system is subject to both exogenous inputs, *d*, which enter the system as filtered disturbances, *w*, as well as control inputs, *u*. The system responds by a measurable system output, *y*, which results in some performance of the system, *z*. A system controller, *K*, is present to supervise the system and make inputs as necessary to ensure system performance conforms to organizational objectives. Many systems can be described using this simple notation, although the exact form of the transfer functions G_i, G_{sys}, and G_o may not always be known. For our purposes here, we will assume that the system operates continuously and the controller, *K*, is an individual human operator. Such a system description might represent an operator controlling an unmanned aircraft system or the officer of the deck standing watch on the bridge of ship. Thus, our problem is to determine the minimum number of individuals that are needed to staff the function, *K*, with the constraint that their predicted task effectiveness must be above some *a priori* threshold. Additionally, it would be desirable, once this minimum number of individuals has been established, to determine how to schedule their duty periods such that their overall average predicted task effectiveness is maximized.

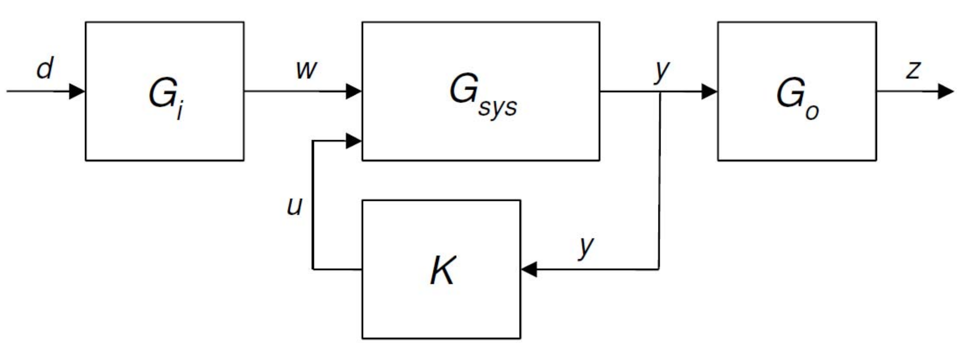


Figure VII-1. Block diagram of a generic dynamic system.

C. THE SLEEP, ACTIVITY, FATIGUE, AND TASK EFFECTIVENESS (SAFTE) MODEL

The SAFTE model is shown emblematically in Figure VII-2 using a system dynamics modeling stock and flow diagram. The conceptual architecture of the SAFTE model centers on a sleep reservoir, representing sleep-dependent processes that govern the capacity to perform cognitive work. Using the language of system dynamics modeling, the stock of this reservoir is cognitive work capacity. Sleep is a replenishing flow into the reservoir, while wakefulness is a depleting flow out of the reservoir. Replenishment, in terms of sleep accumulation, is determined by information about the time-of-day of sleep, reservoir level (i.e., sleep debt), and sleep quality (i.e., sleep fragmentation). The system modeled in Figure VII-2 provides output in terms of performance effectiveness, which is simultaneously modulated by circadian effects and the level of the reservoir (Hursh et al., 2004).

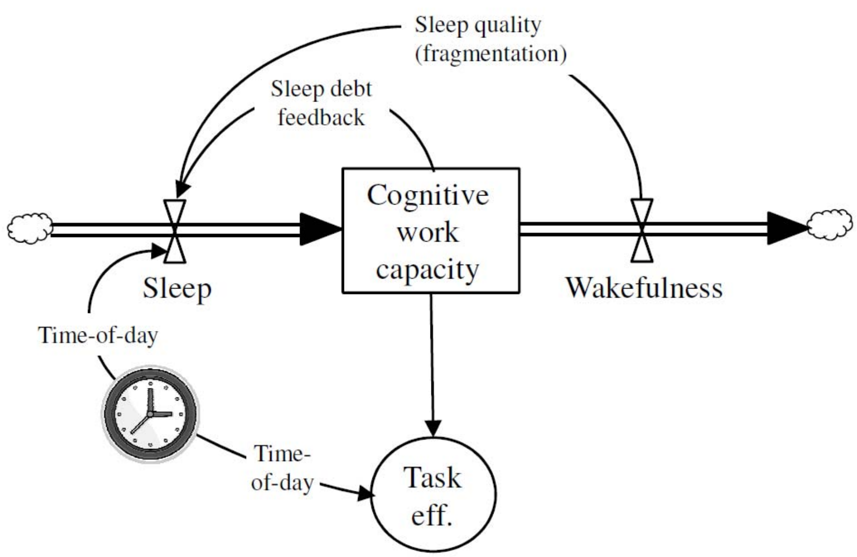


Figure VII-2. Stock and flow diagram of the SAFTE model.

The SAFTE model has been shown to predict changes in cognitive capacity, as measured by standard laboratory tests of cognitive performance, with reported coefficients of determination ranging from 89%–94%. It is presumed these cognitive

tasks measure changes in the fundamental capacity to perform a variety of real-world tasks that rely on such cognitive skills as discrimination, reaction time, mental processing, reasoning, and language comprehension and production. Although specific military tasks may vary in their reliance on these skills, Hursh and colleagues (2004) assert that it is reasonable to assume that changes in military task performance will correlate with changes in the underlying cognitive capacity. Hence, there is an expected monotonic relationship between measured changes in cognitive capacity and military task performance.

Based on the structure of the SAFTE model, the reservoir or stock of cognitive work capability, shown emblematically in Figure VII-2, will remain within some finite range if an individual maintains a constant wake-sleep schedule—that is, the reservoir will exhibit a time-averaged equilibrium state. The stock and flow diagram also shows that sleep accumulation is dependent on information regarding "sleep quality," which is modeled as the contiguity, or conversely, fragmentation of sleep. The software implementation of the SAFTE model (i.e., FAST) addresses sleep quality in terms of the sleep environment and the average number of interruptions to sleep expected in that environment. The FAST software provides the following ordinal scale for describing sleep environments:

- *Excellent*: 0 interruptions per hour
- *Good*: 1-2 interruptions per hour
- *Fair*: 3-5 interruptions per hour
- *Poor*: 6 or more interruptions per hour

These values are equated to 60, 50, 40, and 30 minutes of effective sleep per hour, respectively.

Given the implications of the SAFTE model structure, it is clear that two classes of variables must be considered: schedule and sleep environment. The schedule determines the timing and duration of sleep and wakefulness, and in conjunction with sleep quality, determines the equilibrium state of the reservoir. In principle, the equilibrium state of the reservoir correlates inversely to the degree to which an individual is fatigued, the latter being a direct concern of the survivability domain of Human

Systems Integration (HSI). Likewise, the sleep environment is a determinant of sleep quality, which modulates sleep accumulation, and in turn, the equilibrium state of the reservoir. Since the sleep environment is shaped by the physical environment of sleeping or berthing areas (e.g., adequate space, temperature and lighting control, noise attenuation, etc.), it is a direct consideration of the habitability domain of HSI.

D. AN OPERATIONS RESEARCH PERSPECTIVE

The operations research community focuses on the formulation of mathematical models of complex engineering or management problems and how to analyze them to gain insight about possible solutions. The three fundamental concerns in forming operations research models are the decisions open to decision makers, the constraints limiting decision choices, and the objectives that serve as criteria for rating the relative preference of decision choices. Optimization models, which are also called mathematical programs, are a class of operations research models that represent problem choices as decision variables, which maximize or minimize objective functions of the decision variables subject to constraints on variable values expressing the limits on possible decision choices. Once a problem has been formulated as an optimization model, one can systematically search for optimal solutions, the latter being feasible solutions that achieve objective function values as good as those of any other feasible solution (Rardin, 1998).

Part of the art of constructing mathematical formulations of complex problems is to see past the unique circumstances of the individual problem and recognize general problem types, even if by analogy. The present problem clearly resembles a shift scheduling and staff planning model, where the work is already fixed and we need to plan the resources to accomplish it. The main element in any staff planning model is the covering constraint, which assures that the work periods chosen provide enough worker output to cover requirements over each time period (Rardin, 1998); that is,

$$\sum_{shifts} \left(\text{output/worker}\right) \cdot \left(\text{number on duty}\right) \geq \text{period requirement}.$$

In this case, we express the period requirement in terms of predicted task effectiveness, and we consider shifts in terms of organizationally permissible sleep-wake

cycles. Next, without intending to sound dehumanizing, we contemplate a worker on a shift as being a metaphorical vessel containing a reservoir of cognitive work capacity, such as is depicted in Figure VII-2. For each worker, periods of wakefulness are associated with a discharging flow from the reservoir and periods of sleep are associated with a recharging flow into the reservoir. The output for a worker during a particular period, again expressed in terms of task effectiveness, will be a combined function of the state of their reservoir and their intrinsic diurnal cycle.

If we limit the number of workers on duty during any particular period to unity, we are forced to select a worker from some shift whose predicted task effectiveness meets or exceeds the period requirement for each and every period. Since a decision to use a worker from a particular shift equates to gaining that person in the organization, the objective is simply one of minimizing the number of shifts used to cover all work periods. Solving this staff planning model will yield the manpower optimal solution. However, we may extend this problem one step further by repeating the analysis, but this time restraining ourselves to use no more than the optimal number of shifts and seeking the objective of maximizing the average task effectiveness over all periods. In essence, we are looking for the best arrangement of duty periods given the minimum number of workers. Solving this secondary problem will yield a constrained (in terms of manpower) optimal solution for average task effectiveness. In sum, this is the central logic underlying the optimization programming method described in the following section.

E. THE BASIC MIXED INTEGER LINEAR PROGRAM

The Task Effectiveness Scheduling Tool (TEST) is a modest mixed integer program that assigns persons to wake-sleep cycles and variable duty periods in an attempt to provide coverage of some continuous system function using the minimum quantity of personnel, while simultaneously ensuring individuals exceed an *a priori* predicted task effectiveness criterion during duty periods. The program then ensures that the temporal scheduling of duty periods maximizes average predicted task effectiveness over a 24-hour period. This section presents the formulation of the model with data given in lowercase symbols and decision variables in uppercase symbols.

1. Indices and [Cardinality]

$q \in Q$ — set of ordinal ratings of sleep quality [~4].

$s \in S$ — set of wake-sleep schedules [~72].

$t \in T$ — set of time periods [~48].

2. Data and [Units]

req_eff — required human task effectiveness [%].

$safte\_data^{q}_{s,t}$ — predicted task effectiveness for time period t when following schedule s with sleep quality q [%].

$work\_rule$ — organizational limit on maximum hours of service [periods].

Data on predicted task effectiveness is provided in a matrix with 72 rows and 48 columns. Each row corresponds to a unique schedule, s, consisting of a 6-, 7-, or 8-hour continuous sleep period and a corresponding continuous wake period. Each column corresponds to a time period, $t = 0000, 0030, 0100, \ldots, 2330$, where each t is a 30-minute interval and $t = 0000$ begins at midnight. Wake periods start on a subset of the collection of time periods, $t' \in T$, corresponding to the integer hours of the day, which is to say that $t' = 0000, 0100, 0200, \ldots, 2300$. Thus, S is an exhaustive combinatorial collection of permitted continuous sleep and wake periods. Each schedule, s, in the collection of possible schedules, S, was simulated in FAST version 1.6 over a 30-day period, and the predicted task effectiveness for each time period, t, on the 30th day of the simulation, is recorded in the matrix. Predicted task effectiveness is set to zero during time periods of sleep. Additionally, task effectiveness is set to zero for the 60 minutes prior to and after the sleep period to account for hygiene and other preparatory activities, which would necessarily make an individual unavailable for assignment.

FAST provides for the ability to set an ordinal rating of sleep quality (i.e., excellent, good, fair, or poor) during the sleep period, which impacts the predicted task effectiveness. It is possible to enlarge the matrix of predicted task effectiveness to

consider a quadruplet of schedules, varying in terms of sleep quality, for each primary schedule, *s*, in the collection of possible schedules, *S*: $s = \left\langle s_{excellent}, s_{good}, s_{fair}, s_{poor} \right\rangle \middle| s \in S$. However, this approach adds little to the model, as any attempt to optimize task effectiveness will naturally lead to a choice of $s_{excellent}$ in the absence of some penalty function. Thus, the other elements in the quadruplet will not be selected, but the larger matrix will drive a correspondingly larger decision matrix, and in turn, unnecessarily increase computational burden—a reasonable concern when dealing with integer programs. From a more pragmatic perspective, sleep quality can be ascribed as a function of the environment in which sleep is attempted. Consequently, sleep quality may be fixed *a priori* based on the habitability considerations present within the problem context for which a schedule is being sought.

For the aforementioned reasons, the second approach is used in the subsequent model formulation. Accordingly, separate predicted task effectiveness matrices are developed for each ordinal rating of sleep quality, *q*. The choice of *q* is fixed at q', where $q' \in Q$, and the corresponding predicted task effectiveness matrix, $safte\_data_{s,t}^{q'}$, is incorporated in the model as data. The ensuring sections will suppress further reference to sleep quality for the purpose of economy of notation.

3. Variables

$ASSIGN_{s,t}$ — binary decision variable to assign a person following schedule *s* to cover time period *t*.

$D_{s,t}$ — difference variable used to determine a change in the state (i.e., on or off duty) of a person following schedule *s* at time period *t*.

$MANPOWER_s$ — binary decision variable to utilize a person on schedule *s*.

4. Constraints

(C1) $\sum_s ASSIGN_{s,t} = 1 \ \ \forall \ t.$

(C2) $\sum_t ASSIGN_{s,t} \leq work\_rule \ \ \forall \ s.$

(C3) $\sum_{s} safte\_data_{s,t} ASSIGN_{s,t} \geq req\_eff \quad \forall\ t.$

(C4) $D_{s,t} \geq ASSIGN_{s,t} - ASSIGN_{s,t-1} \quad \forall\ \text{s}, t > 1.$

(C5) $D_{s,t} \geq -ASSIGN_{s,t} + ASSIGN_{s,t-1} \quad \forall\ \text{s}, t > 1.$

(C6) $\sum_{t|t>1} D_{s,t} \leq 2 \quad \forall\ s.$

(C7) $MANPOWER_{s} \geq ASSIGN_{s,t} \quad \forall\ t, s.$

(C8) $ASSIGN_{s,t} \in \{0,1\} \quad \forall\ t, s.$

(C9) $MANPOWER_{s} \in \{0,1\} \quad \forall\ s.$

(C10) $0 \leq D_{s,t} \leq 1 \quad \forall\ s, t > 1.$

5. Objective

Minimize $Z = \sum_{s} MANPOWER_{s}.$

Once the value of the manpower objective is minimized (that is, Z^* is determined), a new constraint is created

(C11) $Z^* \geq \sum_{s} MANPOWER_{s}.$

The program is then solved for the following objective:

$$\text{Maximize } \frac{\sum_{s}\sum_{t} safte\_data_{s,t} ASSIGN_{s,t}}{48}.$$

Solving the first objective establishes the minimum number of persons required to provide coverage of some continuous system function while simultaneously ensuring individuals exceed an *a priori* predicted task effectiveness criterion during duty periods. Solving the second objective seeks to maximize the average predicted task effectiveness of this minimum number of individuals. Thus, the program first establishes the optimal quantity for manpower to satisfice performance requirements, and then it determines how to optimize performance given this now constrained quantity of manpower.

Constraint (C1) is a set partitioning constraint requiring that exactly one person from the collection of wake-sleep schedules, S, belongs to a solution for time period t.

Constraint (C2) tallies the number of time periods an individual on schedule *s* is assigned to provide coverage of a function and enforces organizational hours of service rules. The special case where an organization has no hours of service rules can be simply addressed by setting *work_rule* to 48, which corresponds to the maximum number of time periods in the predicted task effectiveness data matrix. Constraint (C3) enforces the requirement that the predicted task effectiveness of an individual following schedule *s* assigned for duty during time period *t* meets or exceeds some prespecified criterion; alternatively, for each time period, *t*, one could use a filter to only consider the subset of schedules, s', where $s' \in S$, for which predicted task effectiveness meets or exceeds the criterion.

Constraints (C4) and (C5) assess whether a change in assignment status occurs for a person following schedule *s* between time period $t-1$ and period *t*. Constraint (C6) enforces an upper limit on the number of changes in assignment status that can occur for a person following schedule *s*. By setting this limit at two, assigned duty periods are forced to be continuous. This avoids the undesirable result where individuals are assigned to multiple, disjoint time periods. Constraint (C7) acts as a manpower counter: it is set to unity for a person on schedule *s* if they are assigned for any time period, *t*.

Constraints (C8) and (C9) establish the binary decision variables. Constraint (C10) fixes the upper bounds on the nonnegative variable, $D_{s,t}$, at unity.

F. RESULTS AND DISCUSSION

1. Case 1: High Task Effectiveness Criterion

When inadequate attention is paid by system developers to human factors engineering considerations, a potential outcome is “human factors high drivers” (Directorate of Human Performance Integration, n.d.). Such drivers include tasks that require very high levels of sustained human performance, whether that is in terms of vigilance and monitoring, cognitive workload, or physical exertion. This case examines the trade-off between the human factors engineering and manpower domains of HSI that occurs when a requirement is generated that necessitates a high degree of sustained task effectiveness.

Figure VII-3 illustrates the TEST results when the required task effectiveness criterion is set to 95% and sleep quality is assumed to be good—that is, reasonable attention is paid to habitability domain considerations. Each row in the figure corresponds to a single person following a fixed wake-sleep cycle.

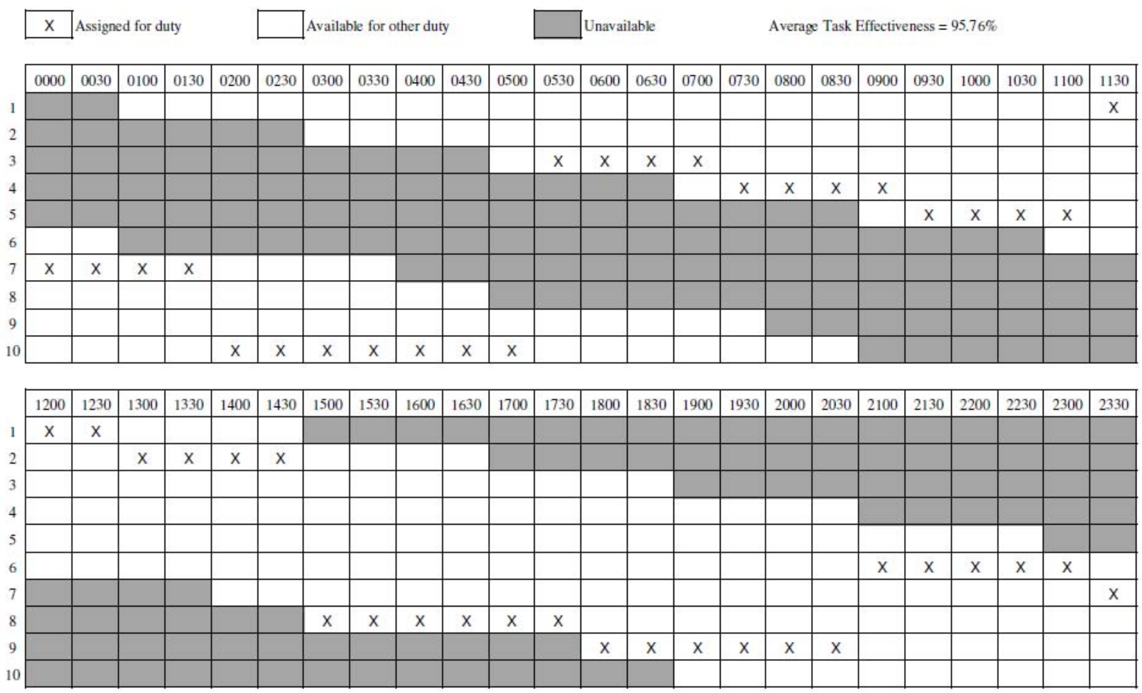
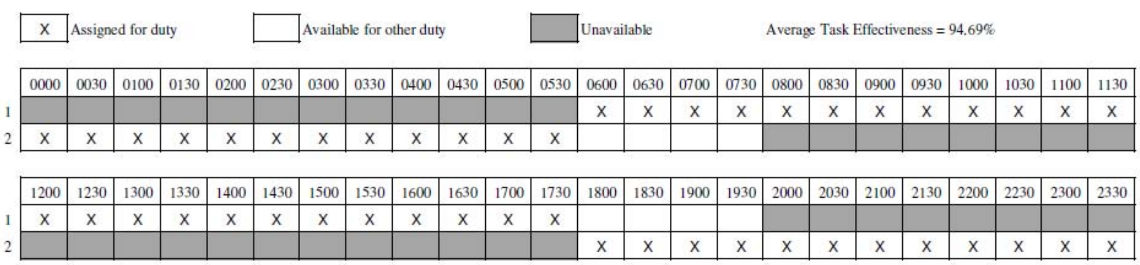


Figure VII-3. Task Effectiveness Scheduling Tool results when the predicted task effectiveness criterion is set to 95% and sleep quality is rated as good.

The shaded boxes in Figure VII-3 are indicative of time periods where a person is unavailable: the first two periods (i.e., one hour) for hygiene and preparatory activities, the next 16 time periods (i.e., eight hours) for sleep, and the last two periods for hygiene and preparatory activities. The nonshaded boxes marked with an “X” are indicative of those time periods when an individual’s predicted task effectiveness meets or exceeds the criterion and they are scheduled to cover the high driver task. The other empty, nonshaded boxes are time periods where a person is available to work, but their predicted task effectiveness in below the criterion. Thus, this time can be allocated to working on less demanding tasks and other personal activities.

What is readily apparent from Figure VII-3 is that human factors high-drivers can lead to excessive manpower requirements—in this case 10 people—to provide sufficient human cognitive resources for the task at hand. Since physiologically based manpower modeling is seldom used in current practice, it is quite likely that individuals charged with developing the system manpower estimate would allocate far fewer than 10 people to cover such a high-driver task. What then results is an unrecognized or implicit trade-off, whereby decreased or more variable performance is accepted, increased systems safety risks are entertained, or both.

Figure VII-4 illustrates the dramatic impact on manpower than can be achieved by mitigating human factors high-drivers during systems development. In this case, the predicted task effectiveness criterion is reduced to 90% and sleep quality is unchanged. While the change in criterion appears relatively modest, the corresponding change in required manpower is dramatic. What previously necessitated 10 people working no greater than 6.5-hour duty periods is now accomplished using only two people working 12-hour shifts.



X Assigned for duty | Available for other duty | Unavailable | Average Task Effectiveness = 94.69%

	0000	0030	0100	0130	0200	0230	0300	0330	0400	0430	0500	0530
1												
2	X	X	X	X	X	X	X	X	X	X	X	X

	0600	0630	0700	0730	0800	0830	0900	0930	1000	1030	1100	1130
1	X	X	X	X	X	X	X	X	X	X	X	X
2												

	1200	1230	1300	1330	1400	1430	1500	1530	1600	1630	1700	1730
1	X	X	X	X	X	X	X	X	X	X	X	X
2												

	1800	1830	1900	1930	2000	2030	2100	2130	2200	2230	2300	2330
1												
2	X	X	X	X	X	X	X	X	X	X	X	X

Figure VII-4. Task Effectiveness Scheduling Tool results when the predicted task effectiveness criterion is set to 90% and sleep quality is rated as good.

2. Case 2: Organizational Hours-of-Work Rules

Sometimes it is the case that individuals performing major system functions belong to professions that are governed by regulatory policies that dictate maximum work periods and minimum rest periods (Miller, Matsangas, & Shattuck, 2007). Often these policies are influenced by nonphysiological considerations such as personnel availability,

mission requirements, and organizational standard operating procedures. Figure VII-5 illustrates the impact of enforcing an hours-of-work rule limiting duty periods to no greater than 10 hours. With the exception of the constraint on hours-of-work, there are no differences in the settings of the model parameters used in the analysis displayed in Figure VII-4 and that shown in Figure VII-5. While the task could be done effectively by two people (Figure VII-4), organizational constraints require that a third person be added to the manpower estimate (Figure VII-5). There is no operationally significant improvement in average predicted task effectiveness (94.69% versus 94.64%) between the two manpower models, but one would expect there to be significant differences in terms of system life-cycle costs. Observations such as this should, at minimum, prompt questions regarding the rationale for the hours-of-work rule.

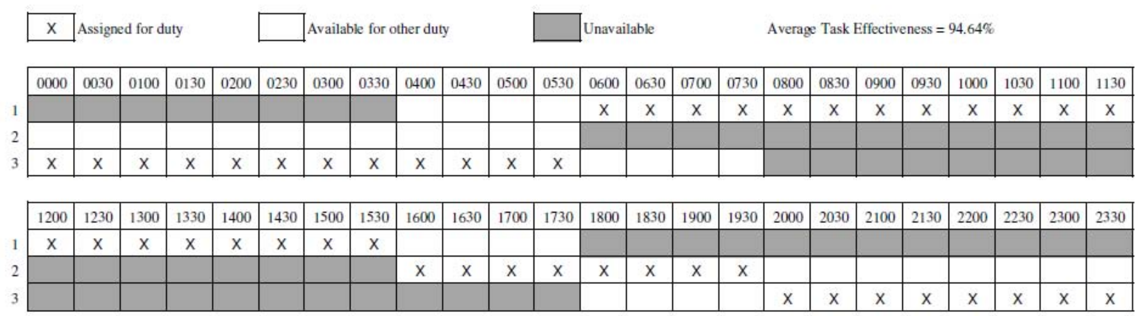


Figure VII-5. Task Effectiveness Scheduling Tool results when the predicted task effectiveness criterion is set to 90%, sleep quality is rated as good, and a 10 hours-of-work rule is enforced.

It is also worth noting in Figure VII-5 that the maximum average task effectiveness is obtained using non-uniform duty periods. The traditional, heuristically based approach to scheduling shift work would lead managers to establish three 8-hour shifts based on the principle of equity (Miller, 2006). In contrast, a physiologically based approach leads to a 10/4/10-hour, 3-shift system. Thus, this case illustrates nicely the disadvantage of using simple scheduling heuristics.

3. Case 3: Sleep Quality

It is generally acknowledged by HSI practitioners and system users that habitability domain considerations are important in sustaining human performance. It is also well recognized by these same individuals that senior decision makers tend to be reluctant to accept or vigorously advocate for system requirements that can be said to be focused on "comfort." Even when such requirements are accepted, they are often the first to be sacrificed when issues of system development cost, schedule, or performance surface.

Figure VII-6 illustrates the case where habitability domain considerations are not given due diligence with regard to their impact on human performance. In this scenario, sleep quality is set at poor and the predicted task effectiveness criterion is relaxed to 77.5%, which corresponds to the threshold for the "criterion line" on the current FAST graphical display. The FAST criterion line equates to the performance of a person following loss of an entire night's sleep. It provides yet another planning heuristic for determining whether a particular schedule is acceptable. However, the validity of this heuristic is certainly questionable, particularly if, for example, a system was designed under the assumption that the operator would perform with a task effectiveness of at least 90%. Nevertheless, even with the reduction in the task effectiveness criterion, it takes eight people—some only suitable for two hours per day—to provide effectual coverage. Contrast this with the observation from the first case that two people can provide more than effectual coverage when sleep quality is set at good. The difference of six individuals between the two scenarios, which can be entirely attributed to the change in the model setting for sleep quality, is quite significant when considered in terms of system life-cycle costs. To summarize, comfort pays!

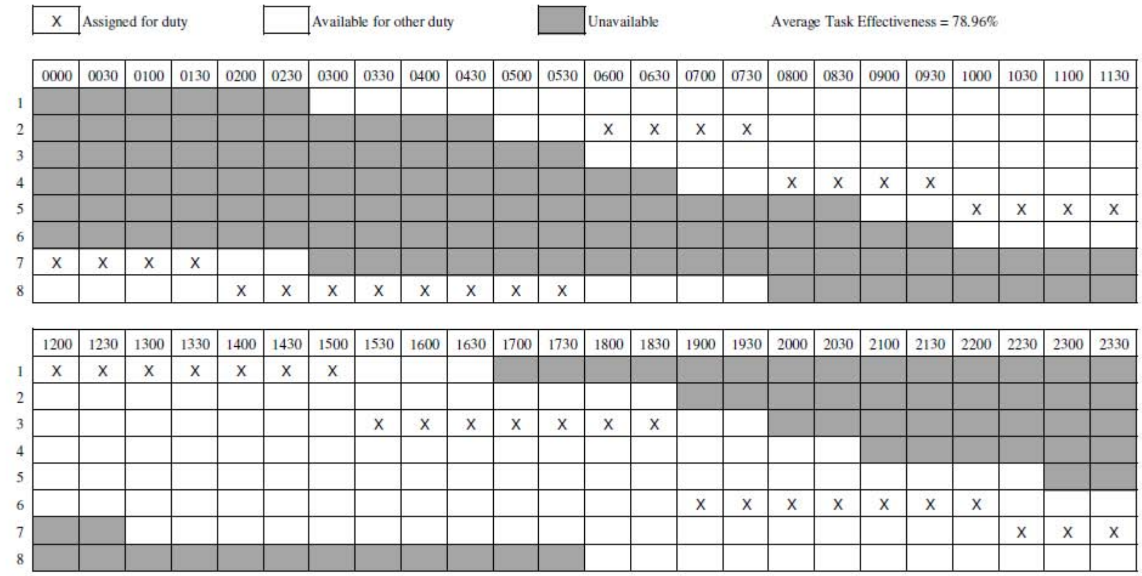


Figure VII-6. Task Effectiveness Scheduling Tool results when the predicted task effectiveness criterion is relaxed to the FAST criterion line of 77.5% and sleep quality is rated as poor.

G. CONCLUSION

In this chapter, we developed a novel approach to staffing and shift schedule planning that offers two key advantages over conventional approaches. First, it allows organizational planners to import data generated from FAST simulations—in essence, the results of individual simulation experiments—into an analytic model whereby answers to the question of optimality can be found by mathematical techniques. Thus, reaching the optimal staffing and shift scheduling solution becomes a less elusive and more deterministic process. Second, it recognizes the inflexible boundary of human capacity and makes explicit the imperative to acknowledge human limitations in the design of staffing and shift schedule solutions. This process should help foster a more holistic approach to designing solutions, thereby taking advantage of the potential trade space that exists between the manpower, survivability, habitability, and human factors engineering domains of HSI. These domains, in turn, involve consideration of issues related to personnel quantity, fatigue (and inversely, the availability of cognitive

resources) and its impact on personnel quality, sleep quality and the opportunity for recovery of cognitive resources, and task demands for cognitive resources, respectively.

By and large, the approach demonstrated here involves nothing uniquely new, either in terms of the biomathematical modeling of fatigue or optimization programming. Rather, it is a new way of using data from biomathematical models of fatigue to systematically find optimal staffing and shift schedule solutions—a way that should be appealing to system developers and force planners. While the model formulation used in this chapter specifically optimizes in terms of manpower, many alternative formulations are possible with minimal modification of the kernel of the model. Similarly, while the model was formulated to address staffing for a single system function (e.g., function *K* in Figure VII-1) requiring a single human controller, it is a simple matter to scale up the model for more complex systems. For instance, incorporating more than one system function would primarily involve the addition of an index set, *f*, to the model formulation where $f = \{K_1, K_2, \ldots, K_n\}$. Likewise, the number of individuals required to simultaneously perform the controller function, *K*, may be easily changed by modifying the right-hand side of the assignment constraint (C1).

To summarize, we expect that coupling biomathematical fatigue models and optimization programming will prove useful in developing physiologically balanced staffing and shift scheduling plans. Further work on this topic should examine the tractability of more complex shift schedule options such as rotating-shift solutions. Additionally, given the potential computational burden of even relatively simple-appearing discrete optimization problems, consideration should be given to the applicability of data filtering, linear programming (LP) relaxations, or both on the analysis of TEST-derived discrete models. Finally, it would be useful to consider how the HSI domain trade-offs that were demonstrated to be inherent in this approach may be incorporated into larger systems analyses.

H. APPENDIX – GAMS CODE

```
Options
     SOLPRINT =     OFF,
     DECIMALS =       2,
```

```
     LIMCOL   =      10,
     LIMROW   =      10,
     RESLIM   =     300,
     ITERLIM  =99999999,
     LP       =   cplex,
     MIP      =   cplex,
     OPTCR    =   0.001 ;

Sets
      t     time periods  / 0000, 0030, 0100, 0130, 0200, 0230, 0300,
                            0330, 0400, 0430, 0500, 0530, 0600, 0630,
                            0700, 0730, 0800, 0830, 0900, 0930, 1000,
                            1030, 1100, 1130, 1200, 1230, 1300, 1330,
                            1400, 1430, 1500, 1530, 1600, 1630, 1700,
                            1730, 1800, 1830, 1900, 1930, 2000, 2030,
                            2100, 2130, 2200, 2230, 2300, 2330 /

      s     schedules     / s600*s623,s700*s723,s800*s823 / ;

Table safte_data(s,t)
$ondelim
$include good_data.csv
$offdelim ;

Scalar
     req_eff     minimum required task effectiveness in percent /90/

     work_rule   maximum allowable continuous hours of work /48/ ;

Variables
      ASSIGN(s,t)
      D(s,t)
      MANPOWER(s)
      OPT_MANPOWER
      OBJ1
      OBJ2 ;

Binary variable ASSIGN ;
Binary variable MANPOWER ;
Positive variable D ;
Positive variable OPT_MANPOWER ;

Equations
     cover(t)            enforce one person assigned for each hour
     length(s)           enforce organizational hours of work rules
     effectiveness(t)    enforce minimum effectiveness requirement
     statechange1(s,t)   check for change state change from period t-1
                         to t
     statechange2(s,t)   check for change state change from period t-1
                         to t
     workcycle(s)        ensure person scheduled at most one work
     convert(s,t)        convert assign decision to manpower
     control(s,t)        control upper limit of D
     objective1          total manpower objective function
     goal_constraint     fix optimal manpower
```

```
     objective2          average reliability objective function ;

cover(t)..  sum(s, ASSIGN(s,t)) =e= 1 ;

length(s)..  sum(t, ASSIGN(s,t)) =l= work_rule ;

effectiveness(t)..  sum(s, safte_data(s,t)*ASSIGN(s,t)) =g= req_eff ;

statechange1(s,t)$(ord(t) gt 1)..  D(s,t) =g= ASSIGN(s,t) - ASSIGN(s,t-
                                   1) ;

statechange2(s,t)$(ord(t) gt 1)..  D(s,t) =g= - ASSIGN(s,t) +
                                   ASSIGN(s,t-1) ;

workcycle(s)..  sum(t$(ord(t) gt 1), D(s,t)) =l= 2 ;

convert(s,t)..   MANPOWER(s) =g= ASSIGN(s,t) ;

control(s,t)..  D(s,t) =l= 1 ;

objective1..  sum(s, MANPOWER(s)) =e= OBJ1 ;

Model fastassign1 / cover, length, effectiveness, statechange1,
                    statechange2, workcycle, convert, control,
                    objective1 / ;

Solve fastassign1 using mip minimizing OBJ1 ;

OPT_MANPOWER.FX = OBJ1.L ;

goal_constraint..  sum(s, MANPOWER(s)) =e= OPT_MANPOWER ;

objective2..  (sum((s,t), ASSIGN(s,t)*safte_data(s,t)))/48 =e= OBJ2 ;

Model fastassign2 / cover, length, effectiveness, statechange1,
                    statechange2, workcycle, convert, control,
                    goal_constraint, objective2 / ;

Solve fastassign2 using mip maximizing OBJ2 ;

File results /results.csv/ ;
results.pw=4096 ;
Put results ;
Put "Average task effectiveness:", OBJ2.L / ;
Put "  " / ;
Put @24
Loop(t,
   Put t.tl:6 ;
   ) ;
Put / ;
Loop(s,
   if( SUM(t,ASSIGN.L(s,t))>0,
     Put "Schedule ", s.tl ;
     Loop(t,
       if( ASSIGN.L(s,t)>0,
```

```
          Put '     X' ;
        else
          Put '      ' ;
        );
      );
      Put /;
    ) ;
);
Putclose
```

I. BIBLIOGRAPHY

Directorate of Human Performance. (n.d.). *Air Force human systems integration handbook: planning and execution of human systems integration*. Brooks City-Base, TX: Author.

Hursh, S. R., Redmond, D. P., Johnson, M. L., Thorne, D. R., Belenky, G., Balkin, T. J., Storm, W. F., Miller, J. C., & Eddy, D. R. (2004). Fatigue models for applied research in warfighting. *Aviation, Space, and Environmental Medicine*, 75(3 Suppl), A44–A53.

Miller, J. C. (2006). *Fundamentals of shiftwork scheduling* (Report No. AFRL-HE-BR-2006-0011). Brooks City-Base, TX: Human Effectiveness Directorate, Air Force Research Laboratory.

Miller, N. L., Matsangas, P., & Shattuck, L. G. (2007). Fatigue and its effect on performance in military environments. In P. A. Hancock & J. L. Szalma (Eds.), *Performance under stress* (chap. 12, pp. 231–249). Burlington, VT: Ashgate Publishing Company.

Neri, D. F. (2004). Fatigue and performance modeling workshop, June 13–14, 2002. *Aviation, Space, and Environmental Medicine*, 75(3 Suppl), A1–A3.

Rardin, R. L. (1998). *Optimization in operations research*. Upper Saddle River, NJ: Prentice Hall.

Shay, J. (1998). Ethical standing for commander self-care: the need for sleep. *Parameters*, 28(2), 93–105.

VIII. LESSONS FROM THE DISCOURSE: WHAT HAVE WE LEARNED?

> The only thing harder than getting a new idea into the military mind is to get the old one out.
>
> — Sir Basil H. Liddell Hart
> *Innovations in the Strategic Air Command*

A. OVERVIEW

This discourse on human systems integration (HSI) consists of ten chapters, each of which is relatively unique in terms of the subject matter addressed and the methodologies employed. Excepting the introductory front matter, we may reasonably assert that each chapter is an independent scholarly work in its own right. This approach is certainly the manner in which the chapters in this discourse were initially constructed, and consequently, each chapter closes with conclusions and recommendations that are specific to the thesis of that chapter. However, these chapters tell a meta-story that reflects both my learning and the evolution of my thinking about HSI during my postgraduate studies. What remains here is to bring closure to that story.

A major purpose in undertaking this discourse was to tackle, head on, two fundamental questions: *What is human systems integration (HSI) and how should one think about HSI problems?* As discussed in Chapter I, the objective in asking this question was to develop a coherent systems method that would improve the integration of the HSI domains to create sustainable systems while preserving consideration of system stakeholder preferences. Addressing this question required that we first put the concept of HSI in some context, both in terms of a philosophy and a Defense Department program.

B. SUMMARY OF CONCEPTS

In Chapter II, "Human Systems Integration Philosophy," we addressed the issue of an HSI philosophy with the implicit assumption that HSI is an emerging discipline. We began by first tracing the origins of HSI philosophy to the early human factors

movement, which approached the problem of human performance in systems from the reductionist perspective embodied by the scientific method. We then discussed the limitations of this approach in dealing with the complexity inherent in the problem, which led to the emergence of the systems-oriented discipline of HSI based on sociotechnical systems theory and the concept of joint optimization of personnel and technological subsystems. We next considered the types of problems HSI typically encounters, often referred to as "messes" or "wicked problems," which entail evolving sets of interlocking issues and constraints that can be managed, but not solved, and for which there are many stakeholders having divergent values, all of whom must be satisfied to some extent. We subsequently considered the role of HSI in addressing these problems by examining its logical placement within a system of systems methodologies. Accordingly, we suggested that various soft systems approaches should first be employed to make problems tractable for HSI in terms of clarity of objectives. In turn, HSI could then be used to make problems more tractable for hard systems approaches by reducing problem complexity.

In contrast, "A Brief History of the Emergence of the Defense Department's HSI Program" (Chapter III) approached the HSI concept based on a historical analysis of HSI as it was programmatically instantiated within the U.S. military. We observed that the idea for an HSI program first emerged in the 1960s as a result of the spread of systems analysis from the Defense Department to the Army, due in large part to efforts to reform the Army's logistics system. Thus, early HSI proponents, such as Weisz, focused on better integrating human factors engineering and operations research to more broadly represent human considerations in weapon system analyses. However, neither wide spread recognition nor high level advocacy for HSI was forthcoming until the Army underwent a doctrinal and organizational renaissance in the late 1970s and early 1980s, driven in large part by fears of an apocalyptic war with the Soviet Union. These conditions led to a rise of science-based military power as the Army sought to leverage high technology to achieve a credible parity with the numerically superior Soviet forces. A crisis ensued during in the early 1980s in the design of the Army of Excellence, caused in large measure by the need to find personnel to create two new light infantry divisions. The Army's solution was to better utilize its personnel resources, especially those tied up

in the maintenance and support of increasingly complex, highly technological weapon systems. In 1983, a coalition of senior military and civilian leaders began an HSI discourse, including the materiel acquisitions and personnel communities, which was eventually to be institutionalized in the Army's bureaucracy as MANPRINT. In the ensuing decade of the 1990s, this discourse was carried over to the Defense Department bureaucracy where it became formally codified as HSI.

The lesson learned from the juxtaposition of these two conceptual views of HSI (i.e., philosophy versus program) was the *rejecting of the notion that HSI is simply "post-modern" human factors*. HSI as a philosophy evolved within the context of the larger systems movement that occurred in the 1960s in response to the issue of irreducible complexity. HSI emerged in response to real-world, macroergonomic political and military considerations that resulted in an organizational crisis. This crisis, in the simplest of terms, was caused by technological complexity and its effects on personnel. Thus, the fundamental impetus for HSI was *complexity*.

Allowing philosophy to inform method, the lessons learned from the historical analysis were used to characterize and illustrate an approach to addressing HSI considerations early in a weapon system acquisition process. The following prime directive—the highest level of abstract, objective statement of purpose—was proposed for an HSI program: *To produce sustained system performance that is humanly, technologically, and economically feasible.* Based on an analysis of this prime directive, and with an implicit reference to sociotechnical systems theory, the following definition of HSI was derived:

> A philosophy applied to personnel and technological subsystems within organizations in pursuit of their joint optimization in terms of maximally satisfying organizational objectives at minimum life cycle cost. Its practice is concerned with the specification and design for reliability, availability, and maintainability of both the personnel and technological subsystems over their envisioned life cycle.

We then asserted that the principle approach to HSI should involve the integration of the behavioral sciences, human factors engineering, and operations research to more broadly represent human considerations early in weapon system analyses and in the products that evolve from these analyses.

Accomplishing the HSI prime directive necessitates a holistic perspective of the performance and economic trade space formed by the synthesis of the HSI domains, and as a result, the consideration of individual domain interventions in terms of tradeoff decisions. Accordingly, in Chapter IV, "The Human Systems Integration Trade Space Problem," we took up the primacy of tradeoffs in HSI. We expanded our conceptualization of HSI to consider both a macro-HSI and micro-HSI trade space. Macro-HSI focuses on the development and utilization of human resources within organizations that own and operate technological systems that are, in turn, the subject matter of micro-HSI; macro-HSI is concerned mainly with macroergonomic considerations of organizational and work-system design. In contrast, micro-HSI concentrates on individual technological systems and subsystems and, at least in its contemporary implementation, is strongly oriented towards human factors engineering or microergonomic considerations. Thus, an overarching goal of HSI must be one of making tradeoffs that are organizationally net positive to avoid creating future problems. Such a goal is tractable if macro-HSI considerations are used to first bound or constrain the micro-HSI trade space. Then one may deliberatively consider the micro-HSI trade space in the systems decision process by integrating Simon's research strategy of efficient multifactor design of experiments, Kennedy and Jones' isoperformance approach, and coupling isoperformance with utility analysis through means such as physical programming.

Although domain tradeoffs are a central element of HSI, there are very few studies that aptly illustrate the integration of the behavioral sciences and human factors engineering with the tools and methodologies of operations research. The purpose of the subsequent three case studies was to fill that void. To grasp tradeoffs in any meaningful manner, the basic models developed in these case studies were necessarily abstractions or simplifications of reality to delineate clearly the basic domain relationships and

interactions. Therein, probably, lies one of their strongest merits, particularly for the HSI novice. With respect to these analyses, I claim no infallibility, particularly where practical analysis required judgments to be made to move past inevitable impasses. Hopefully, though, these analyses further illuminated the contexts within which judgments about HSI tradeoffs can take place. Chapters V, VI, and VII each presented three case studies that illustrated the use of different data sources: a preexisting opportunistic dataset of potential Air Force unmanned aircraft pilots, a prospective dataset of Army Soldiers in Basic Combat Training, and data derived from simulation of staffing and shift scheduling solutions using a biomathematical model.

C. CASE STUDY REVIEWS

In Chapter V, "Isoreliability Models for Human Systems Integration Domain Tradeoffs—Choosing a Personnel Supply Source for Future Unmanned Aircraft System Operators," we considered how to address the reliability of the personnel subsystem within the joint optimization problem described by sociotechnical systems theory. We utilized an opportunistic dataset from an Air Force study evaluating the impact of prior flight experience on acquisition of unmanned aircraft system (UAS) operator skills. Based on a derivation of the isoperformance methodology, we developed a simple logistic regression-based analysis for relating the personnel and training domains of HSI to the proportion of proficient UAS operators, which allowed us to express human performance probabilistically in terms of *isoreliability*. We also demonstrated the feasibility of both including logical decision variables in isoperformance models and incorporating such models into larger discrete optimization models that were then used to analyze aggregated system functions. This analysis established the potential to integrate isoreliability models into the systems engineering process, thereby allowing consideration of personnel and training domain tradeoffs in terms of total system reliability—and indirectly, in terms of systems safety.

The study described in Chapter VI, "Human Systems Integration Domain Tradeoffs in Non-Technical Systems—Improving Soldier Basic Combat Training," provided an excellent illustration of the application of HSI outside of the context of

systems engineering and program management. Recognizing that adolescents comprise the majority of military accessions, this study evaluated the impact of alterations in sleeping and waking patterns on measures of Soldier performance and other indicators of individual functioning during basic combat training. We conducted the study using the behavioral sciences paradigm, and thus employed an experimental methodology and multi-variable statistical techniques drawn from experimental psychology. The results indicated that, even after controlling for factors contributing to individual differences, adjusting the scheduled sleep period in a phase delayed direction (i.e., later bedtime and wake-up) was associated with increased daily total sleep and modest improvements in some indicators of daytime functioning. We then transitioned from the behavioral sciences to the HSI paradigm by reformulating a subset of the study hypotheses in terms of mathematical tradeoff functions, thereby making possible their direct incorporation into the "system analytic thinking process." Specifically, using the isoperformance methodology, we constructed tradeoff models for both rifle marksmanship performance and occupational health in terms of the personnel and survivability domains of HSI.

In the final case study, "Human Systems Integration Domain Tradeoffs in Optimized Manning—The Task Effectiveness Scheduling Tool" (Chapter VII), we investigated an approach for determining optimal schedules in terms of the timing of sleep-wake periods and the assignment of performance sensitive duties when given an *a priori* task effectiveness threshold—essentially an issue of human operational availability $\left(A_o^h\right)$. The necessary data were generated from numerous simulations conducted using the Defense Department's Fatigue Avoidance Scheduling Tool, which is based on the validated Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE) model. We then constructed the Task Effectiveness Scheduling Tool (TEST), a mixed integer program that assigns persons to wake-sleep cycles and variable duty periods to provide coverage of a system function using the minimum quantity of personnel while simultaneously ensuring individuals exceed a specified task effectiveness criterion during duty periods. The program then ensures that the temporal scheduling of duty periods maximizes averaged predicted task effectiveness over a 24-hour period. By exercising TEST with several use cases, we observed that the mathematical program facilitates explicit

exploration of the trade space that exists between the manpower, survivability, habitability, and human factors engineering domains of HSI.

D. EPILOGUE

We began the introduction to this discourse with the observation that there is much ambiguity about what exactly HSI is and how it should work. This discourse was motivated by the wish—and really, my need—to fill this void. There was also a fair sense of urgency in bringing this discourse to fruition given my forthcoming assignment within the U.S. Air Force's still nascent HSI work force. A quick survey of the current environment in which Air Force HSI finds itself suggests, perhaps alarmingly so, that time is not working in our favor. We need to begin to deliver on the promises made by the U.S. Air Force Scientific Advisory Board in their 2004 report, *Human System Integration in Air Force Weapon Systems Development and Acquisition* (Report No. SAB-TR-04-04). Otherwise, we risk, and should probably expect, a serious loss in advocacy and support for a robust Air Force HSI program in the immediate future.

While there will always be those who believe we should solve the problems of HSI by primarily developing better HSI tools, the reality is that we need to move forward with a more pragmatic approach, grounded in science and/or the lessons of history, which can achieve some modicum of success in today's defense acquisition programs given the available tools at hand. Fortunately, guided by our historical insights into HSI, we can reasonably assert that the necessary tools, by and large, already exist in the form of the experimental and statistical methodologies of the behavioral sciences and human factors engineering and the tools and techniques of operations research. As was aptly illustrated by the three case studies presented in this discourse, we can handily accomplish HSI tradeoff analyses today, but multiple techniques are needed to formulate and conduct these analyses—and hence the futility of the search for a single, comprehensive HSI tool. Consequently, there is a clear need for are creative, thinking HSI practitioners who are sufficiently savvy in the behavioral science, human factors engineering, and operations research such that they can utilize a portfolio of techniques and credibly participate on study teams conducting systems analyses.

Some may sensibly question whether, given Popper's criticisms of historicism, we have gone down a proverbial rabbit hole to emerge in an academic wonderland. Certainly the accounting of HSI provided in this discourse differs markedly from those provided by others in the field over the last decade. While it is impossible to ascertain with complete certainty that we have avoided the potential pitfalls of historicism, a recent experience appears to suggest that we have, indeed, reached a sensible accounting of HSI. To briefly describe that experience, having recently completed the last of the case studies in the discourse, I was invited to attend a 3-day training course on pre-acquisition analyses that was sponsored by the Air Force HSI community and was lead by the Air Force's Office of Aerospace Studies. The intent of the training course was to discuss the implications of the Weapons Systems Acquisition Reform Act of 2009, particularly with regards to the need for more comprehensive systems analyses prior to Milestone A. Although originally advertised as a training course, the majority of the time was spent in an interactive and lively discussion—or what might be better described as a joint iterative learning cycle (SSM from Chapter II) – between those in the studies and analyses community and we in the Air Force HSI community over "where HSI fits in" pre-acquisition systems analyses. Reassuringly, the major themes that emerged confirmed both the need identified by and the primary thrusts put forth in this discourse. I would argue that this observation provides some tentative corroboration in support of the overarching thesis put forth in this discourse. Definitive proof of this thesis, however, will need to await future attempts to actually put it into practice through participation on pre-acquisition study teams—an objective for my next military assignment within the U.S. Air Force.

In conclusion, the efforts in this discourse served to accomplish two things: 1) extract the lessons learned from a historical analysis of the emergence of HSI both as a philosophy and as a Defense Department program, and 2) use those lessons to characterize and illustrate a mathematical and technical approach to addressing HSI considerations early in the weapon system acquisition process. While by no means a novel approach, for reasons unbeknownst to me it has not been used by those seeking to advance the field of HSI. Nevertheless, we clearly observed that the general systems

discourse that occurred over the latter half of the last century, coupled with pressing organizational factors within the U.S. Army, were the principal forces that shaped and drove the emergence and formal recognition of HSI. While we primarily considered the trajectory of HSI through the Army and on to the Defense Department, future work is warranted to develop equivalent case histories for the evolution of HSI within the U.S. Air Force and Navy—and other government agencies as appropriate. Such case histories should prove a rich source for high-level lessons regarding the influence of organizational context on the implementation of HSI that would be very relevant to those responsible for Service-specific HSI programs. For as Edmund Burke (1729–1797), the British statesman and philosopher, is credited with famously warning, “Those who don’t know history are destined to repeat it.”

THIS PAGE INTENTIONALLY LEFT BLANK

IX. A NEW HUMAN SYSTEMS INTEGRATION

Think like an architect—not like a bricklayer (Warden & Russell, 2002, p. 179).

A. WHAT IS WRONG WITH CLASSIC HUMAN SYSTEMS INTEGRATION?

1. Attitudes and Understanding

Quite a lot is wrong, and if you think otherwise, then try discussing human systems integration (HSI) as a discipline with many of its current practitioners. Invariably there are two reactions: they are too busy to think about such things or they vent frustration about a lack of clarity of purpose / underlying theory / evidence of success. Many practitioners do not believe that HSI is a separate discipline. Instead, they prefer to think of it in terms of human factors, systems engineering, or plain common sense—although there is plenty of evidence to suggest that such sense is not common, at least in the Defense Department.

Many practitioners describe HSI as an attitude or state of mind. Such words as "holistic" and "human-centered design" will often emerge to encapsulate the "human > hardware ≈ software" view that most feel intuitively sets HSI apart in some degree. Behind such slogans, however, is often little more than a self-rationalizing, if not pleasing, belief that keeping a laser-like focus on humans throughout the system acquisition process is both necessary and sufficient to yield some form of "goodness," the latter often described in vague terms of enhanced systems performance or decreased operating costs. The more rigorous descriptions of HSI, including those contained in the few books dedicated to the subject, appear to be based on the premise that the human element in complex systems can be addressed through a set of ad-hoc processes and actions—many oriented to the individual domains of HSI—and often accompanied by only a modicum of analytical results. Nearly all current HSI practitioners would probably not ascribe to the idea that HSI is a hybrid discipline involving the behavioral sciences and operations research—an idea that was established in Chapter III.

If the run-of-the-mill HSI practitioner does not believe in HSI as discipline, then there is indeed a problem. While HSI has not had many spectacular successes, high profile problems, such as the shortfalls in the design of the Air Force's Predator ground control station, can be seen to have occurred because the simplest of HSI practices were not observed—in that case, the Joint Program Office postponed addressing issues related to training, maintainability, human factors, and reliability of the system until after the prototype system was transitioned to the eventual operator (i.e., the Air Force) (Thirtle, Johnson, & Birkier, 1997).

2. Right Approach, Wrong Result?

An interesting aspect of classic HSI is the resistance to it—or at least the proclivity to minimize it—by program managers and budgeters, particularly when faced with programmatic schedule or cost constraints. Classic HSI strongly emphasizes front end analysis and human factors engineering activities as those are the types of activities that are within the purview of program managers and systems engineers working in the weapon system acquisition process. Other HSI domain considerations are then addressed primarily in terms of their relation to the human factors engineering domain (Pew & Mavor, 2007). Program managers, seeing no comprehensive framework to trade off current materiel development considerations and future non-materiel considerations, become frustrated by promises of significant returns on investment from human factors engineering activities. Stoking this frustration is the fact that those benefits will be primarily realized by others far removed from the weapon system acquisition process. Even when program managers are supportive of classic HSI, compelling success stories remain rare.

The U.S. Army's RAH-66 Comanche acquisition program is perhaps the most notorious and best-documented example of this last point. The Comanche was the first major Army program to both implement classic HSI considerations into the front-end analysis phase of the materiel acquisition process and to include HSI in the source selection document. Thus, Comanche became a true experimental program, testing where it was possible to introduce advanced technology without creating problems of

unsatisfactory total system performance or increasing personnel demands. Even opponents of the Comanche program were impressed by the advances relative to the standard of normal acquisition practices; it was estimated that the potential cost avoidance in the Comanche program in terms of manpower, personnel, training, and safety was $3.3 billion, equating to an 8,000 percent return on investment for the portion of the program's research and development budget that was attributable to HSI (Booher & Minninger, 2003; Skelton, 1997). Nonetheless, the Army made the decision in 2004 to cancel the Comanche program, though not for reasons related to HSI. To date, no similar Defense Department case study has demonstrated a comparable level of success.

B. A NEW HUMAN SYSTEMS INTEGRATION METHOD

Previous chapters have introduced ideas of philosophy and the issues of emergence, hierarchy, and irreducible complexity; history and the general systems discourse that occurred over the latter half of the last century; the trade space formed by the synthesis of the HSI domains; and many others, which contributed to a new look at HSI. This new look was based on the HSI Hypothesis presented in Chapter II; the objective was to use this Hypothesis as a base upon which to develop concepts and design a methodology for managing HSI issues (particularly early) in the Defense Department's weapon system acquisition process. Since the HSI Hypothesis addresses any and all systems, such an HSI method should be applicable to any system, be it human activity, technological, or any other.

1. Design Guidelines

The design guidelines encapsulating the New Human Systems Integration are shown in Table IX-1.

Table IX-1. New Human Systems Integration guidelines.

Step 1	Establish SOI* objectives and requirements by reference to containing system(s)
Step 2	Identify containing system(s)' strategic human resources objectives
Step 3	Identify sibling systems (vis-à-vis shared human resources) and their interactions that will be perturbed by the SOI
Step 4	Develop SOI design trade space to complement sibling systems in contributing to containing system(s)' objectives
Step 5	Functionally partition SOI and describe required (emergent) human-system performance in terms of response surfaces that are functions of the domains of HSI
Step 6	Reduce response surfaces to isoperformance (tradeoff) equations for incorporation in system analyses
Step 7	Seek a balanced design (joint optimization) that satisfices SOI objectives and requirements
Step 8	Continuously reassess and rebalance the design throughout the life of the SOI

*SOI = System-of-interest.

2. Step 1: Establish SOI Objectives and Requirements by Reference to Containing System(s)

The only tangible value of any system is to be seen in its contribution to its containing system(s)' objectives—see Chapters II and IV. Note the plurality of containing systems; it is quite rare in the real world for any system-of-interest (SOI) to have only one containing system. For example, while weapon systems are contained within some user combatant command, they are also contained within a service's personnel system, training system, logistics system, etc. Thus, reference to relevant containing systems—comprising both materiel and nonmaterial aspects of the SOI—is necessary if one is to successfully seek a balanced design for the SOI.

3. Step 2: Identify Containing System(s)' Strategic Human Resources Objectives

Historically, human (personnel) resources have been the primary driver of system life-cycle costs. Additionally, human resources are often a constrained resource, particularly within the Defense Department where Congress legislates arbitrary manpower ceilings. Clearly, if an organization is to manage its largest cost driver, it will need a human resources investment strategy. As described in Chapter IV, a primary means for implementing such a human resources investment strategy in the weapon system acquisition process is to specify "system proactive" manpower, personnel, and training constraints. System proactive constraints are deliberately formulated to shape the design of future systems so that aggregate demand for human resources within the containing system(s), both in terms of personnel quantity and quality, is driven toward some explicit set of human resources investment goals. In essence, rather than defining what the target audience will need to be given the demands of the technological system, one asks what the design of the technological system should be so that some idealized future target audience is sufficient. In economic terms, the emphasis is on human resources demand management rather than supply management (i.e., recruiting). Of course, any human resources investment strategy must be sensitive to macroergonomic considerations such as labor relations. Nonetheless, the possibility of implementing such a human resources investment strategy presupposes the existence of some underlying information technology architecture to support aggregation, analysis, and forecasting of the organizational supply and demand for human resources.

4. Step 3: Identify Sibling Systems (Vis-à-vis Shared Human Resources) and Their Interactions That Will Be Perturbed by the SOI

Within the containing system(s) are sibling systems with which the SOI will interact and which will be disturbed by that new interaction. Recall from the discussion of sociotechnical systems theory in Chapter II that every system, including the SOI, is comprised of personnel and technological subsystems (Figure IX-1).

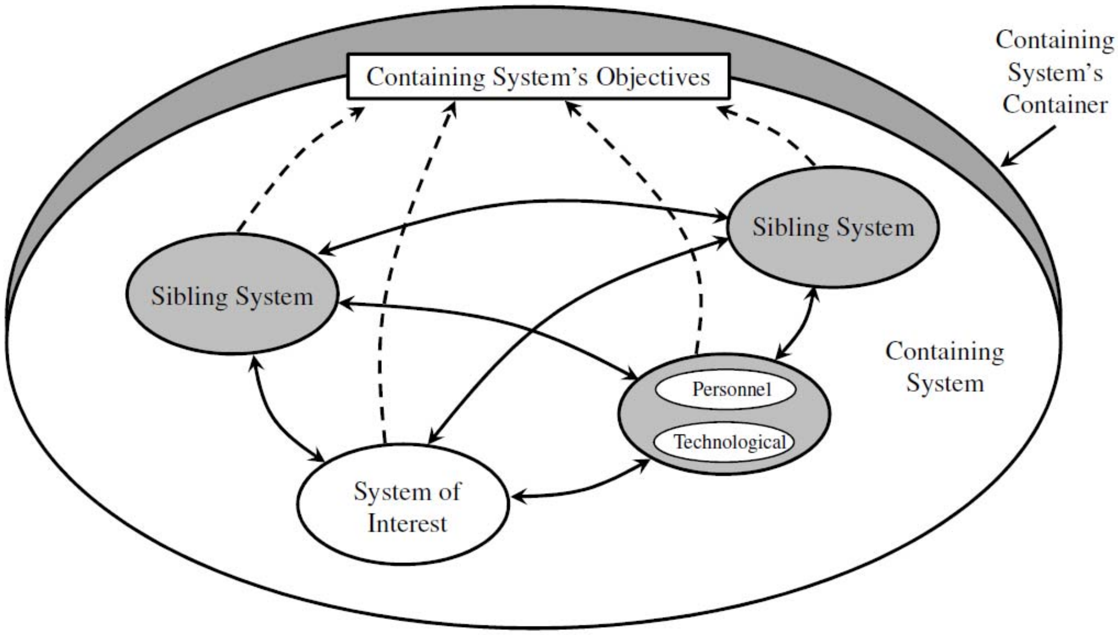


Figure IX-1. A family of interacting systems, to include a system-of-interest and its sibling systems, all existing within the environment provided by the containing system.

Each of these personnel subsystems is also part of a larger organizational human (personnel) resources system, which is constrained in terms of its total size and composition. Thus, the sibling systems will, in turn, interact, thereby changing the environment within the containing system(s), and hence, the interactions with the SOI (Figure IX-2). In complex situations, such interactions may produce results that are difficult to predict and could potentially be very undesirable. In the case of the Defense Department, the key insight is that many diverse weapon systems may potentially be sibling systems by virtue of shared human (personnel) resources, whether in terms of operations, maintenance, or support.

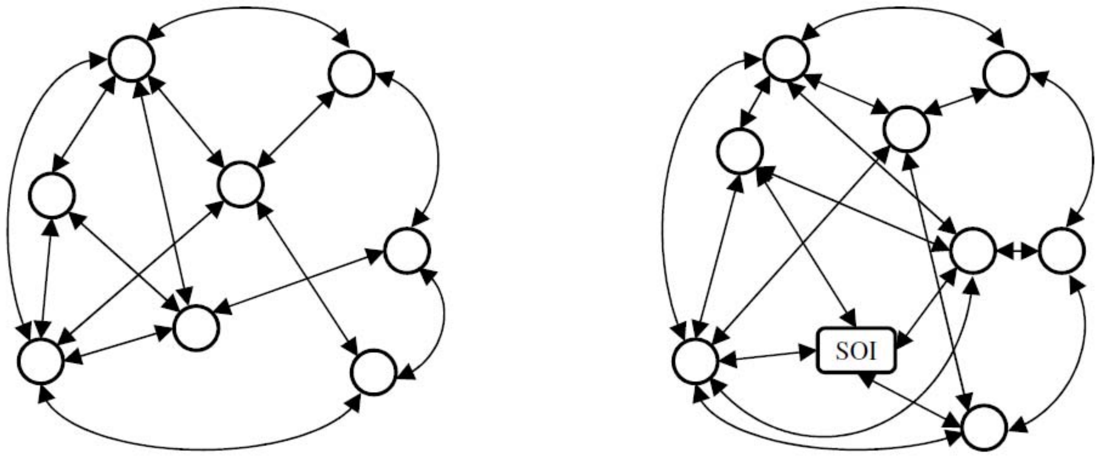


Figure IX-2. Interacting systems readjusting to the addition of a new system-of-interest [From Hitchins, 1992].

5. Step 4: Develop SOI Design Trade Space to Complement Sibling Systems in Contributing to Containing System(s)' Objectives

Attempts to optimize the personnel subsystem of the SOI in isolation may create changes that propagate throughout the environment of the containing system, disrupting the joint optimization of the personnel and technological subsystems comprising sibling systems, and degrading sibling systems' performance. Thus, the value of a SOI's contribution to the containing system is only relevant in the context of the similar contributions of its sibling systems—hence the need to focus on net positive contributions. These considerations throw the most serious doubt on our ability to formulate sensible HSI system requirement specifications by simply considering a new system. Rather, HSI specifications must flow down from an aggregate analysis of human resources relative to all sibling systems. Such specifications then serve to provide the macro-level bounds on the SOI design trade space. However, as was discussed with regards to a human resources investment strategy, the ability to accomplish this step requires the existence of an information infrastructure to allow aggregation, analysis, and forecasting of organizational supply and demand for human resources.

6. Step 5: Functionally Partition SOI and Describe Required (Emergent) Human-System Performance in Terms of Response Surfaces That Are Functions of the Domains of HSI

Designing the SOI invokes many of the classic systems engineering activities, but always from an outward-looking perspective aligned towards the containing system(s). Partitioning, or the process of functional decomposition, is crucial to the successful design of any SOI. It determines the SOI's subsystems and their interactions and interfaces, and of particular interest here, it leads to the allocation of functions between the personnel and technological subsystems. The challenge for the systems practitioner is then to describe the human performance trade space for potential system concepts (i.e., architectures) in terms of the domains of HSI. Recalling the discussion in Chapter II of emergent phenomenon, hierarchy, and the problem of irreducible complexity, describing this human performance trade space necessitates a systemic (i.e., holistic) and systematic approach that heretofore has been absent in traditional, reductionist approaches to human factors.

Describing the human-system performance trade space for the SOI means mapping an emergent property (i.e., human performance) for all feasible settings of its producer elements (i.e., the domains of HSI). Recall that Systems Age thinking is concerned with producer-product rather than cause-effect relationships (Chapter II). As was suggested in Chapter IV, this task of mapping the trade space is best accomplished using the Simonian approach (Simon, 1977) with its emphasis on economical multifactor design of experiments. Based on a program of research marked by progressive iteration, the Simonian approach first employs fractional factorial screening experiments to identify the primary "HSI drivers" (as determined based on Pareto analyses of effect sizes), followed by advanced experimental designs to describe the response surface that emerges from their interactions. These experiments may utilize human-in-the-loop assessments (e.g., virtual environments), modeling and simulation (e.g., IMPRINT), or some combination thereof. Nonetheless, the overall objective of this progressive iteration is to describe performance as a polynomial regression function of its many potential determinants—a format that should be both intuitive and useful to engineers.

7. Step 6: Reduce Response Surfaces to Isoperformance (Tradeoff) Equations for Incorporation in System Analyses

The Simonian approach solves the methodological problem of developing functional models of performance where there are many determinants of potential importance. Jones and Kennedy (Jones, Kennedy, & Stanney, 2004) provide the conceptual approach for reformulating such models into quantitative tradeoff functions that can then be incorporated into weapon system analyses. In so doing, the conceptual divide between behavioral science and operations research was bridged as was first called for by Weisz in his premonitions of the Army's MANPRINT program—see Chapter III. Based on the functional partitioning of the SOI, classical systems engineering provides criterion levels of performance for each function (i.e., functional measures of performance), often stated in the Defense Department in terms of threshold (i.e., lower bound) and objective (i.e., upper bound) values. Given such a criterion level of performance and a desired assurance (or confidence) level, the isoperformance method fixes the dependent variable in the functional model of performance and solves the resulting polynomial regression equation in terms of just the determinants—*á la* the tradeoff function.

8. Step 7: Seek a Balanced Design (Joint Optimization) that Satisfices SOI Objectives and Requirements

It is of no use trying to optimize the parts of the SOI and then join them together—that is system integration, not system design. Nor is it any better to review the overall results of such integration and then attempt to intuitively adjust some of the design parameters. The primary design perspective must be that of a single system, comprised of the combination of personnel and technological subsystems, which is open to the environment and the sibling systems it contains. Consequently, the joint optimization of the SOI's personnel and technological subsystems, with reference to more than one containing system, will be an inherently complex problem.

Formulating a balanced design for the SOI means formulating a mathematical model of what is essentially a complex engineering and management problem, thereby facilitating analysis and insight into possible solutions—that is, employing the methods

of operations research. In terms of HSI, this means coupling isoperformance with utility analysis in terms of a system analytic meta-model (Chapter IV). Such a meta-model should aggregate the individual isoperformance models, created as a result of the functional partitioning of the SOI, into one or more measures of total system performance (e.g., rolling up functional isoreliability models using reliability block diagrams to create a system-level reliability estimate—see Chapter V). In a modification of Jones and Kennedy's approach, the dependent variable in each isoperformance model should be allowed to vary across the range from threshold to objective values rather than being fixed at a criterion level of performance. Next, appropriate bounds for decision variables should be stipulated based on consideration of human resource investment goals (Step 2) and the need to complement sibling systems (Step 4). Lastly, all variables that reflect concerns of containing system(s) of the SOI need to be addressed in terms of utility functions, preferably using physical programming to minimize controversy over the choice of weights and to take advantage of the one-versus-others rule—the latter being an important attribute of physical programming that explicitly works to help ensure a balanced system design.

The modified Jones-Kennedy-Simon approach described above should be familiar to engineers. A similar approach has been used in classic systems engineering to develop what here would be characterized as the technological subsystem of the SOI—see the paper by de Weck and Jones (2006) describing the application of isoperformance to both the design of a satellite and the performance of a sports team. Thus, seeking a balanced design for the SOI needs to lead to a systems analysis that includes isoperformance models for both the personnel and technological subsystems, thereby ensuring that whatever is being optimized can be credibly claimed to involve the joint (or collective) optimization of the personnel and technology subsystems. Moreover, this approach will decrease the likelihood for dominance by either subsystem in the chosen design solution because of the appeals of either technology or human-centered design zealots.

9. Step 8: Continuously Reassess and Rebalance the Design Throughout the Life of the SOI

Given the concept of joint causation (Chapter II), systems practitioners must be concerned with changes in the SOI's environment and corresponding adaptive changes to its subsystems. In all likelihood, joint optimization of the SOI at a particular time will be relatively short lived. Since the technological subsystem, once designed, is relatively fixed, any adaptation that the containing system(s) permits the SOI will fall primarily to the personnel subsystem for implementation. There is a need, then, to manage the evolution of the SOI throughout its life cycle—hence, the characterization of HSI problems as messes or "wicked" problems (again see Chapter II). However, for the most part, much of the original system analysis work should remain valid; the important distinction in subsequent analyses is that those variables that were associated with the technological system in the original analysis will now need to be considered as fixed. This resultant loss of degrees of freedom relative to the original problem highlights the challenge of continuously rebalancing the overall design of the SOI over its life cycle. Moreover, as the problem space becomes progressively more constrained by the loss of decision variables, the point may be reached where the new optimal solution is significantly worse than that obtained in the past.

C. CHALLENGES FOR NEW HUMAN SYSTEMS INTEGRATION

Several challenges spring to mind when looking at New HSI. Foremost is the question of how to operationalize the New HSI design guidelines within the Defense Department. The New HSI approach requires that behavioral scientists and human factors engineers, schooled in traditional reductionist science, be reoriented to understand Systems Age thinking and the related issues of emergence and hierarchy. This means that experiments examining a handful of potential determinants and geared to testing hypotheses need to give way to evolutionary programs of research aimed at mapping multi-dimensional human-system performance response surfaces. Given the marginal utility of the human factors literature in this regard (see discussion of Simon's work in Chapter IV), deliberate programs of research will need to be undertaken to generate these performance response surfaces—an ideal applied research task for the military service

laboratories. While such programs of research could be managed by behavioral scientists and human factors engineers, retrained to be Systems Age thinkers, it will likely be easier to have these programs managed instead by systems oriented and systems thinking New HSI practitioners.

But what then is a New HSI practitioner? If we ascribe to HSI as a discipline, then such a person should be characterized as a professional within this discipline, and thus a skilled practitioner or expert. For the vast majority of practitioners, New HSI will be a learned profession, meaning that there will be the need for preparatory education. As was clearly demonstrated in this discourse, New HSI requires one be a systems practitioner—and consequently a Systems Age thinker—who is sufficiently knowledgeable in systems engineering, the behavioral sciences, and operations research such that they can integrate these fields, reflecting the fact that New HSI is a hybrid discipline. The key to achieving such integration rests with the ability of the New HSI practitioner to take considerations of the behavioral sciences and formulate them for inclusion in mathematical models of complex engineering and management problems, which in turn can be analyzed to gain insights about possible solutions. It is predominately this step of problem formulation, best described as an amalgamation of both art and science, which drives the need for preparatory graduate level education in New HSI. While there are many software tools to solve mathematical models once formulated, at present, problem formulation remains a job for the gray-matter computer comprising the human brain. Clearly then, educating New HSI practitioners will not be quick or cheap—but then again, as the old adage goes, “you get what you pay for.”

Segueing now to the issue of HSI modeling and simulation tools, our discussion of New HSI practitioners should cause us to pause before launching into prescriptions for HSI tools. There currently exist multiple HSI tool repositories, containing literally hundreds of tools, as well as numerous ongoing research and development efforts to improve existing or develop additional HSI tools. One cannot help but wonder if the unspecified shortfall in existing tools is really one of the tools themselves or the inability of classic HSI practitioners to properly formulate problems such that they can then be answered using existing tools. Classic HSI practitioners tend to view tools as a means for

directly solving system design problems. In contrast, New HSI practitioners should understand HSI tools to be primarily sources of data; solving a systems design problem, however, involves correctly formulating a mathematical model of the problem that can then be exercised using the data from the tool to gain insight—an operations research problem. The latter perspective was illustrated by the third discourse case study introducing the Task Effectiveness Scheduling Tool (Chapter VII), which in effect was a mathematical program that utilized data generated from a validated biomathematical simulation—an existing HSI tool—to systemically and systematically explore staffing and shift scheduling solutions (i.e., design of the personnel subsystem). Thus, future work on HSI tools should assess the adequacy of existing models and simulations in generating the data necessary for response surface mapping—Step 5 in the New HSI design guidelines. Further investments in HSI tools should then focus on closing any identified capability gaps in this regard. In the end, however, it must continuously be emphasized that it is the educated New HSI practitioner, and not the tools *per se*, that primarily enable systems to be designed that promise to be flexible, adaptable, reliable, inexpensive to own, and long-lived.

Lastly, from where should New HSI be managed? At least with regards to the Defense Department, Steps 3–5 in the New HSI design guidelines suggest some higher integration authority than that which currently exists for classic HSI—the latter being the program manager for the SOI. Clearly, the approach taken in classic HSI does not readily allow for consideration of tradeoffs with sibling systems. Nor does it provide sufficient emphasis for the continued management of joint optimization of the SOI after its deployment and fielding. Rather than making specific prescriptions for any particular organization, we instead turn to Hitchins (1992) generic reference model and its necessary and sufficient set of internal functions: mission, viability, and resource management. According to Hitchins, mission describes the system purpose, viability establishes the system to pursue that purpose, and resources are used both in the pursuit of the mission and in the maintenance of viability. Together, mission, viability, and resource management are referred to as the management set. Figure IX-3 depicts New HSI as emerging from the synthesis of the functions comprising Hitchins' management

set. Systems thinkers should be able to conceive of a recursive hierarchy of management sets corresponding to the SOI, sibling systems, and containing systems. Consequently, human activity systems (recall Checkland's systems typology described in Chapter II) responsible for New HSI should be established at appropriate levels within an organization's hierarchy of systems where intersections of Hitchins' management functions occur. As applied to the Defense Department, such an approach should yield a substantially different organizational structure for managing HSI than is observable in practice today.

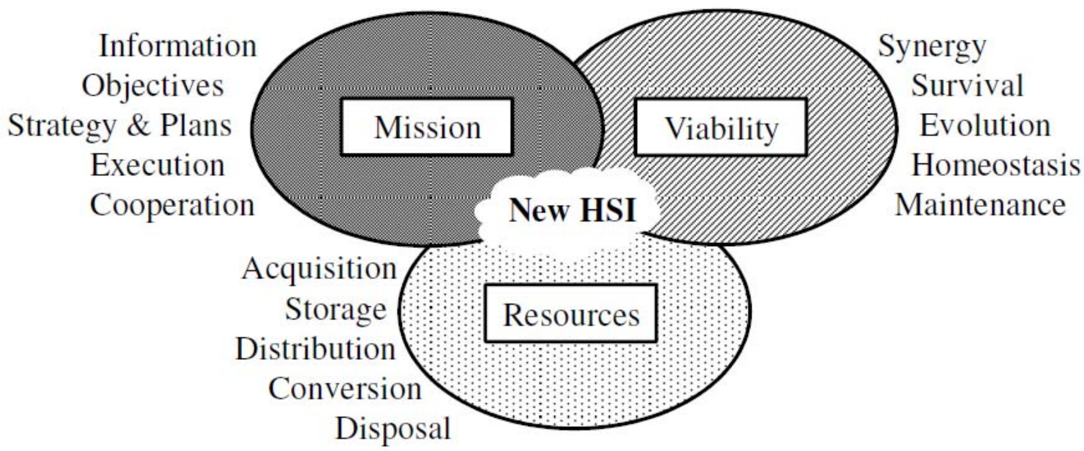


Figure IX-3. Generic reference model and New HSI [After Hitchins, 1992].

D. CONCLUSION

The New Human Systems Integration introduces an approach, unlike classic HSI, based on theory, philosophy, and the lessons of history. It enables specification of system designs with clear purpose, which fit into their environments, and that contribute, along with their siblings, to the objectives of their containing systems. The New HSI does not negate classic HSI—it enhances it, particularly by emphasizing the initial (i.e., pre-acquisition) and final (i.e., post-acquisition) phases of design synthesis, thereby drawing attention to the resource allocation decisions that potentially most impact the emergent properties valued by stakeholders in the containing systems. Such notions as

“top-down requirements” are given substance, since the top can be clearly identified with the objectives of the containing systems, to specifically include the human (personnel) resources system. Classic HSI considerations are retained, but with the essential difference that all systems—the SOI, sibling systems, and containing systems—are considered at all times in the joint optimization of the SOI. Thus, it is possible to design and implement systems that are humanly, technologically, and economically feasible and sustainable. Systems designed using this approach promise to be more balanced and less likely to be dominated by either personnel or technological considerations, which should make them more adaptive, flexible, and resilient over their life cycle.

E. BIBLIOGRAPHY

Booher, H. R., & Minninger, J. (2003). Human systems integration in Army systems acquisition. In H. R. Booher (Ed.), *Handbook of human systems integration* (Chap. 18, pp. 663−698). New York, NY: John Wiley and Sons.

de Weck, O. L., & Jones, M. B. (2006). Isoperformance: analysis and design of complex systems with desired outcomes. *Systems Engineering*, 9(1), 45−61.

Hitchins, D. K. (1992). *Putting systems to work.* Chichester, England: John Wiley and Sons.

Jones, M. B., Kennedy, R. S., & Stanney, K. M. (2004). Towards systematic control of cybersickness. *Presence*, 13, 589–600.

Pew, R. W., & Mavor, A. S. (2007). *Human-system integration in the system development process: a new look*. Washington, DC: National Academies Press.

Simon, C. W. (1977). *New research paradigm for applied experimental psychology: a systems approach* (Report No. CWS-04-77). Washington, DC: Air Force Office of Scientific Research.

Skelton, I. (1997, October 1). *MANPRINT for the U.S. Army*. Statement by Congressman Ike Skelton. Washington, DC.

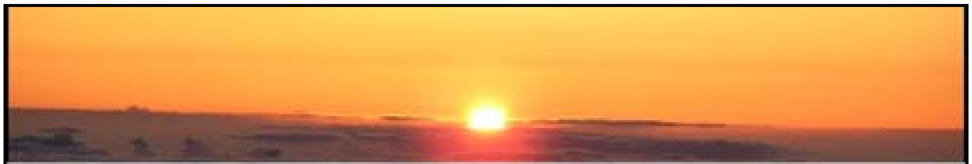
Thirtle, M. R., Johnson, R. V., & Birkier, J. L. (1997). *The Predator ACTD: a case study for transition planning to the formal acquisition process*. Santa Monica, CA: RAND.

Warden, J. A., III, & Russell, L. A. (2002). *Winning in fast time*. Montgomery, AL: Venturist.

THIS PAGE INTENTIONALLY LEFT BLANK

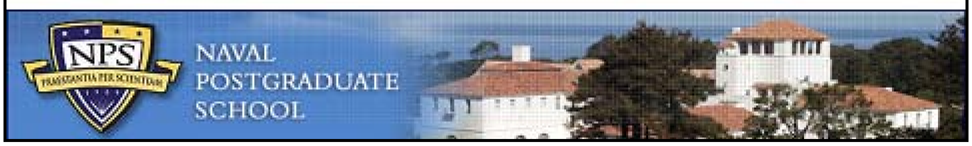
X. APPENDIX: A DISCOURSE IN HUMAN SYSTEMS INTEGRATION BRIEFING SLIDES

The purpose of this discourse on human systems integration (HSI) was to address the central questions of what is HSI and how should one go about thinking about HSI problems. The responses to these questions were developed in terms of a meta-story that spans seven chapters, excluding the introductory front and summary end matter in the discourse. Given the length of the discourse, a first draft summary presentation was prepared for the dissertation committee members. The following slides were those presented to the committee members prior to the dissertation defense; from these, 66 slides were used during the actual dissertation defense.



A Discourse on Human Systems Integration

Anthony P. Tvaryanas



Central Issue:

What is human systems integration (HSI) and how should one think about HSI problems?

HSI Theology:

Human element in complex systems can be addressed through a set of ad-hoc processes and actions, often accompanied by a modicum of analytical results

Today's HSI model:

Redacting the Theology:

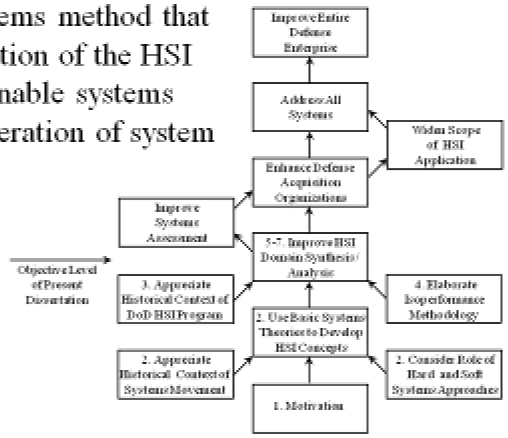
HSI involves integration of ≥ 7 domains $\rightarrow$ HSI solutions described in terms of sets of ≥ 7 domains

<u>Proposition:</u> Accommodating the human element in technological systems is an N-dimensional creativity problem, not a 3-dimensional ergonomics problem

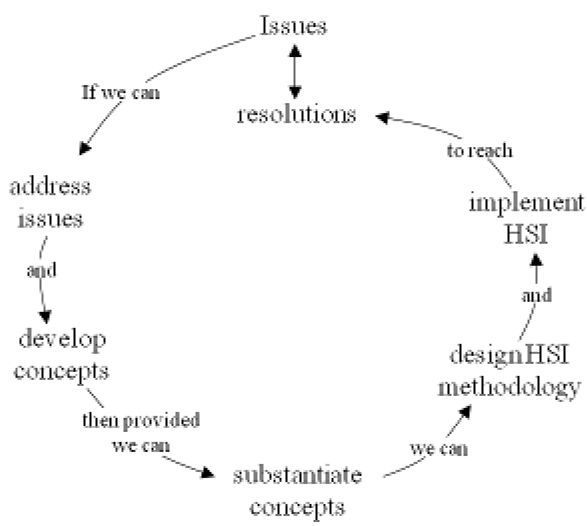
Need a <u>systemic</u> and <u>systematic</u> approach to HSI

Intent Structure:

Mission statement: To develop a conceptually sound systems method that will improve the integration of the HSI domains to create sustainable systems while preserving consideration of system stakeholder preferences.



Approach:



Human Systems Integration Philosophy

2. Systems Context

Problem Statement:

No general consensus on HSI (process, science, professional discipline, etc).

So what is it?

Human Performance in Systems:

Formal scientific discipline focused on humans in systems emerged in WWII (Chapanis et al., 1947)

- First public discussion of this subject given in series of 10 lectures (*Lectures on Men and Machines*) to NPS engineering students (1947)

Problem of human performance in systems parsed in distinct fields of inquiry (Chapanis et al., 1947, Kennedy et al., 1989)

- Psychophysical systems research (time-and-motion engineering & experimental psychology) → human factors engineering
- Personnel psychology → personnel selection
- Educational psychology → training and education

Emergence of "large systems" approaches in 1980s (Kleiner, 2008)

- Macroergonomics
- Human systems integration

Problem of Complexity:

Scientific method attempts to deal with complexity by deconstructing phenomenon into separate parts for study

Practitioners of science manage complexity by dividing knowledge of world into arbitrary subjects or disciplines

Comte's doctrine (1865)

- Evolution of sciences: theological, metaphysical, positive
- Natural order of sciences: physics, chemistry, biology, psychology, social sciences

Checkland (1981)

- Elements of science: reductionism, repeatability, & refutation
- Irreducible complexity → hierarchy of sciences
- Each level of hierarchy characterized by own autonomous problems

Major problem for method of science

Sociotechnical Systems Theory:

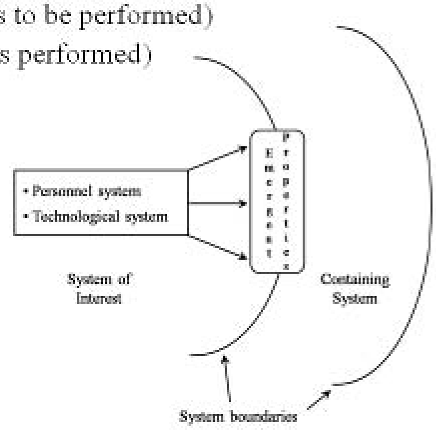
Organizations open systems engaged in process of transforming inputs into desired outcomes

Organizations bring 2 critical factors to bear on transformation process:

- Technological subsystem (tasks to be performed)
- Personnel subsystem (way tasks performed)

Joint causation

Joint optimization



Transition from Machine to Systems Age:

Machine Age: understanding the whole is the sum of understanding its parts (Descartes)

- Reductionism / analysis
- Mechanistic cause-effect relationships
- Deterministic, input oriented worldview

Systems Age: the whole is more than the sum of its parts (Aristotle)

- Expansionism & teleonomy / synthesis
- Probabilistic producer-product relationships
- Stochastic, output/outcome oriented worldview

(Ackoff, 1981)

Systems Thinking:

Gestalt & holism

Emergence & hierarchy

Systems typologies (Boulding, 1956; Jordan, 1968; Checkland, 1981)

Complex adaptive systems (self-organization, emergence, & evolution)

Wicked problems (puzzles, problems, & messes)

Hard & soft systems approaches:

Hard	*Soft*
Substantive rationality	Procedural rationality
• Select from a set of alternative COAs	• COAs must be discovered
• Data available to predict consequences of COAs	• Solutions developed by resolving conflicts over ends & means
• Criterion exists for choosing COA (optimality)	• Satisfice rather than seek optimality

Total Systems Intervention (TSI):

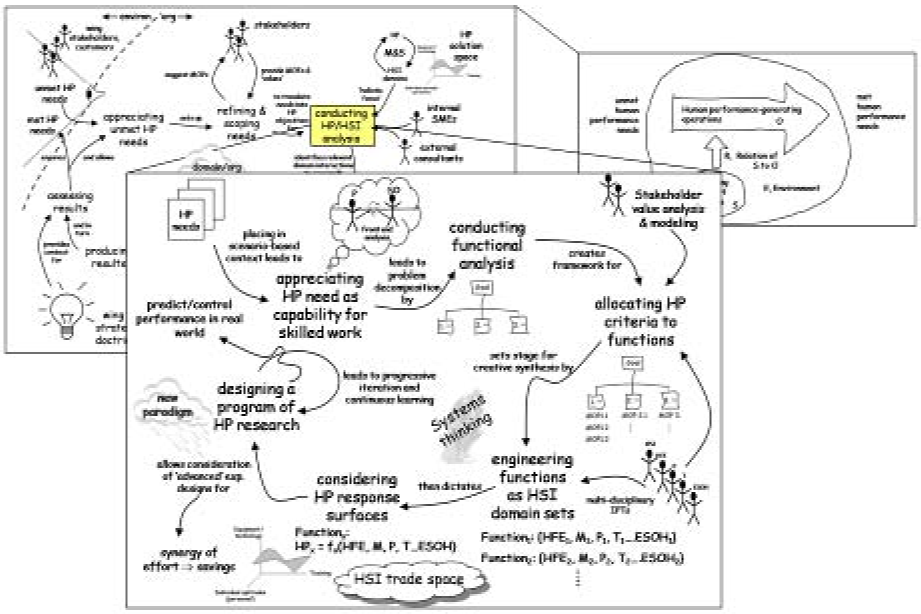
Flood & Jackson (1991) attempt to resolve hard-soft dichotomy → refocus on holistic intent of systems perspective

Meta-methodology focused on creatively surfacing issues organization faces & choosing methods that tackle issues most effectively

System of systems methodologies (SOSM):

	Unitary	**Pluralist**	**Coercive**
Simple	• Operations research • Systems analysis • Systems engineering • System dynamics	• Social systems design • Strategic assumption surfacing and testing	• Critical systems heuristics
Complex	• Viable system diagnosis • General system theory • Socio-technical systems thinking — HSI • Contingency thinking	• Interactive planning • Soft systems methodology	? ? ?

Understanding HSI Through TSI:



Conclusions / HSI Hypothesis:

HSI not "post-modern" human factors → evolution of human factors within context of larger systems movement in response to issue of irreducible complexity

DoD HSI system <u>prime directive</u>:

To produce sustained system performance that is *humanly*, *technologically*, and *economically* feasible

HSI <u>definition</u> (as derived from PD):

HSI is a managerial philosophy applied to personnel and technological subsystems within organizations in pursuit of their joint optimization in terms of maximally satisfying organizational objectives at minimum life cycle cost. Its practice is concerned with the specification and design for reliability, availability, and maintainability of both the personnel and technological subsystems over their envisioned life cycle.

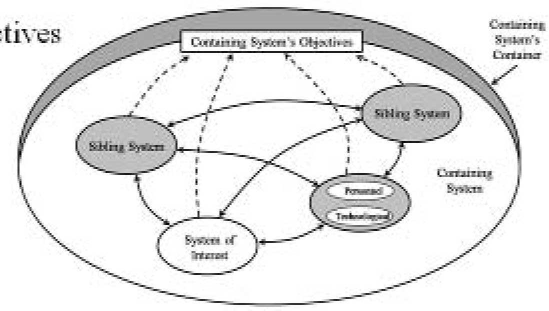
HSI Hypothesis – Insights:

Management philosophy

Must continuously address sustained performance of SOI over its life

Focus on designing for operational feasibility (SOI performs as intended in effective & efficient manner)

Local & global perspectives



Brief History of the Emergence of the Defense Department's Human Systems Integration Program

3. Historical Context

Problem Statement:

Most HSI definitions traceable to DoDI 5000.2 (Deal, 2007)

So what was the genesis of the DoDI 5000.2 description of HSI?

Macroergonomic Issues (1945–1991):

American post-WWII political culture (Edwards, 1988, 1996)

- Premise: Technological choices ↔ political context
- Key elements of U.S. political culture
  - Apocalyptic struggle with former U.S.S.R.
  - Long history of anti-militarist sentiment in American politics
  - Rise of technology-based military power

Postwar geopolitical concerns of U.S. as world power shaped strategic discourse centered on high technology

Technological determinism (Holley, 2004)

- Defined as thesis that superior arms favor victory
- Not true unless superior institutional weapon system acquisition practices yield innovative technology wed to thoughtful doctrine

Rebuilding the Army (1970-1980):

Army concentration on infantry-airmobile warfare during Vietnam

Emerging threats:

- Numerically & qualitatively superior Warsaw Pact forces (1970s)
- Tempo & lethality of Arab-Israeli War (1973)

Reactive period of structural modernization & doctrinal reform focused on armor warfare in Europe

- Gen DePuy (1973–1977) → Division Restructuring Study & doctrine of Active Defense (FM–100 5)
- Gen Starry (1977–1981) → Army 86 Studies & AirLand Battle doctrine
  - Emphasis on technology to counter Warsaw Pact numerical superiority
  - "Big 5" weapon systems: M-1, M-2, AH-64A, UH-60A, Patriot

(Romjue, 1993)

Organizational Crisis – The Army of Excellence:

Division 86 (heavy division) approved by CSA Gen Meyer (1979)

Soviet invasion of Afghanistan & Iran hostage crisis (1979) → need for rapidly deployable light divisions (global defense mission)

- Infantry Division 86 Study (TRADOC)
- 9th ID (Lt Gen Elton) → High Technology Test Bed (HTTB)

Threat & not Army end-strength constraints drove Division 86 design → not affordable with given manpower (1981)

CSA Gen Wickham orders new design: Army of Excellence (1983)

- Create 2 light infantry divisions (7th & 25th ID) & preserve strength in heavy divisions
- Manpower for infantry divisions from combat support → RMA

(Wild, 1987; Dupay, 1988; Romjue, 1993)

Military Reform Movement:

Main Debate:

Reformers

- Overemphasis on technology driving costs out of control
- High technology introduces level of complexity that diminishes readiness
- High technology pushed in areas irrelevant to success in combat (may endanger users)
- Added increment of performance obtained by high technology does not justify cost
- High technology stretches acquisition → critical delays & unexplained technical problems

Technologists

- Technology acts as a force multiplier
- Technology provides force flexibility
- Technology has potential to improve cost & equipment RMA
- Technology is indispensible given the alternatives

(Herzog, 1994)

Spinney Report (1983):

INCREASING WEAPONS COMPLEXITY REDUCES COMBAT READINESS

- Degrades combat skills by causing inadequate and unrealistic training.
- Increases reliability and maintainability problems.
- Increases cost of maintenance.
- Increases dependence on large vulnerable support base.
- Increases economic inefficiency of plans.
- Slows modernization by increasing development/procurement lead times.
- Multiplies magnitude and likelihood of disaster.
- Increases vulnerability to countermeasures.
- Cuts forces, supplies, and munitions to inadequate numbers.

QUESTION

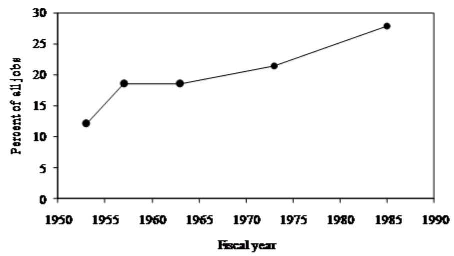
Do the distinctive characteristics generated by weapons *complexity* compensate for these negative qualities?

Effects of Technology of Manpower:

Effects of technology on workforce determined by degree of system *complexity* (Binkin, 1986)

Early 1980s, experience of Army bears out reformers argument:

- Soldiers not realizing predicted performance
- Increased soldier-to-system ratios and soldier skill requirements (Blackwood & Riviello, 1994)



Growth in technical jobs in U.S. Army (Binkin, 1986)

Concurrent Evolution of MANPRINT Concept:

Battle of France (Guilmartin & Jacobowitz, 1985)
WWII aviation (Flanagan, 1954)
Army HEL incorporated in AMC (1962)
Brown Board Study (1967)
Weisz HEL publications (1967–1969)
AR602-1 published (1968)
Army's "lost decade" for materiel development
DARCOM Pamphlet 706-102 (1979)
Kerwin–Blanchard Study (1980)
GAO Weapon System Design Report (1981)
SMI Requirements ("Complexity") Study (1981)
Reverse Engineering Project (1982)
Three Army Science Board Summer Studies (1981–1983)

Army Reg 602-1 (1976):

SYSTEM DESIGN FOR PERSONNEL – MATERIEL INTERFACE					
Medical	Environmental Factors	Personnel-Material Interface	Training	Personnel Requirements	Organizational Factors
• Health • Physical Selection	• Tolerance • Safety • Protection • Performance Enhancement	• Human Capabilities/ Limitations • Materiel Design • Personnel Performance Reliability	• Methods • Media • Equipment • Simulation • Resource Requirements	• Personnel Classification and Selection • Manpower and Career Management • Task Performance • Skill Qualification Testing	• Organization Changes • Management and Organizational Task Analysis • Integrated Logistics Support Program

PERSONNEL-MACHINE-MISSION PERFORMANCE
IN
SYSTEMS DEVELOPMENT AND OPERATIONS

Army MANPRINT Program:

Trifecta of change agents in summer 1983:

- General Maxwell Thurman (VCSA)
- Lt. General Robert Elton (DCSPER)
- Major William Blackwood (DCSPER Studies & Analyses Office)

Window of opportunity for organizational change:

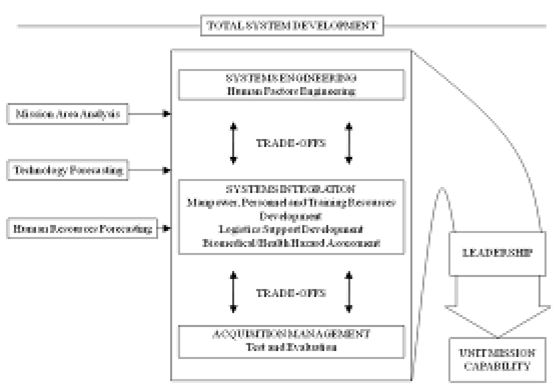
- Elton tasks ODCSPER to develop plan giving personnel community "sense of place and purpose" in WSAP
- Goal to improve manpower & personnel utilization in Army

DSCPER sponsors Army Science Board 1984 Summer Study (*Leading and Manning Army 21*) to start organizational change process

Army Science Board Summer Study 1984:

Recommendations:

1) Single HMPT authority equal to materiel in WSAP
2) Soldier research to improve total system performance
3) HMPT initiatives with staying power in Army organizations & processes



Army MANPRINT Program:

Gen Thurman assigns DA staff responsibility for HMPT to DCSPER (June, 1984)

Multiple parallel initiatives during summer 1984:

- Gen Thompson launches "MANPRINT" program in Army Materiel Command → HFE Task Force to address HMPT in AMC
- TRADOC MPT steering committee
- Health Services Command, Army Safety Center, OTEA, others…

ODCSPER suspicious of initiatives (short-term focus) → gains approval for own plan from GO Steering Committee (Dec 1984)

- Long-term (10-yr) strategy → freeze organizational change

DSCPER Plan:

Primary goal (middle) sandwiched by more acceptable goals

- Improve human performance → improve total system performance
- Improve manpower & personnel utilization in Army at large
- Weapons that are easier to use, maintain & support → improve unit effectiveness & readiness

Implementation plan focus areas:

- Policy & procedures (AR 602-2)
- Marketing & communications (Elton)
- Training & education (Hay Systems)
- Resources
- Research & studies
- Evaluations & applications

Critical intervention points in WSAP:

- Request for proposal
- Source selection
- Test & Evaluation
- Army Systems Acquisition Review Council

Optimizing Performance:

M1 crews score more kills than M60 crews of similar aptitude

Performance of M1 crews less sensitive to aptitude

∴ Army can relax requirements → improve personnel utilization

AFQT category of gunner/tank commander	*Tank equivalent kills*		
	M60	*M1*	*Percent improvement*
I (above average)	10.23	12.75	25
II (above average)	9.51	12.47	31
IIIA (average)	8.52	12.05	41
IIIB (average)	7.47	11.57	55
IV (below average)	5.84	10.72	84

(Binkin, 1986)

Evolution of DoD-level HSI (1987-1991):

Change agents:

- LTC Blackwood: ODSCPER → USD/AQ strategic planning office (1987)
  - Thomas Christie (Boyd Acolyte)
  - Lt Col Michael Pearce (ASD/FM&P)
- Mr. Spurlock (ASA/M&RA) → Weisz champion, proponent for MER

ASD/FM&P signals commitment to MANPRINT/HSI goals

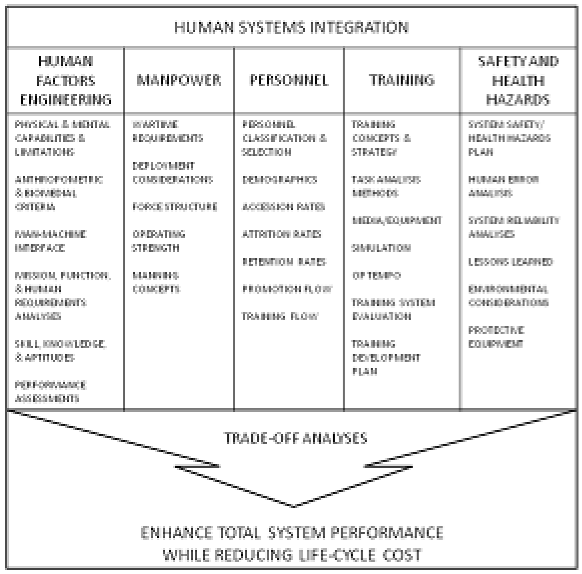
- Sponsors DoDD 5000.53 (1988) requiring MER & MPTS criteria in WSAP
- Established HSI office (Lt Col Pearce)

ASD/FM&P focuses on MPT → weak advocacy leads to demise of HSI office with departure of Pearce

DoDD 5000.53 incorporated into 1991 revision of DoDI 5000.2 (AQ)

- HSI formally appears in name, enhanced in definition & scope
- Diminished content in subsequent DoDI 5000.2 revisions

DoDI 5000.2 (1991):



Conclusions:

HSI concept emerged as result of spread of systems analysis from RAND to DoD → integration of HFE and OR to more broadly represent human considerations in weapon systems analyses (1960s)

Soviet threat in 1970s and 1980s drove Army doctrinal & organizational renaissance → rise of science-based military power (technology) → complexity crisis centered on personnel domain

HSI concept rediscovered as means to resolve crisis by improving "tooth-to-tail" ratios for weapon systems (1983)

Systems discourse involving materiel acquisition & personnel communities → institutionalization of MANPRINT in Army (1984), later HSI in DoD (1991)

The Human Systems Integration Trade Space Problem

4. Isoperformance

Problem Statement:

Weisz's paradigm necessitates that HSI considerations be formulated for inclusion in weapon systems analyses.

So, how do we model HSI tradeoffs in systems analyses?

HSI Conceptual Models:

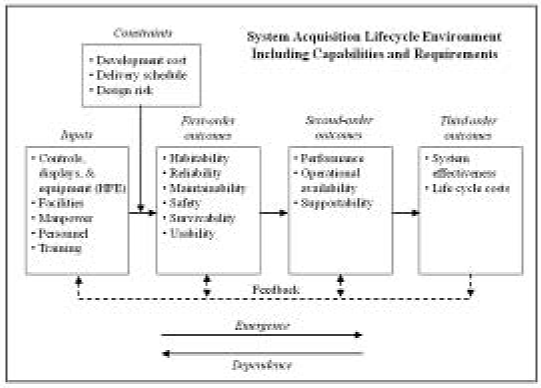
Analysis & Decomposition

Synthesis & Integration

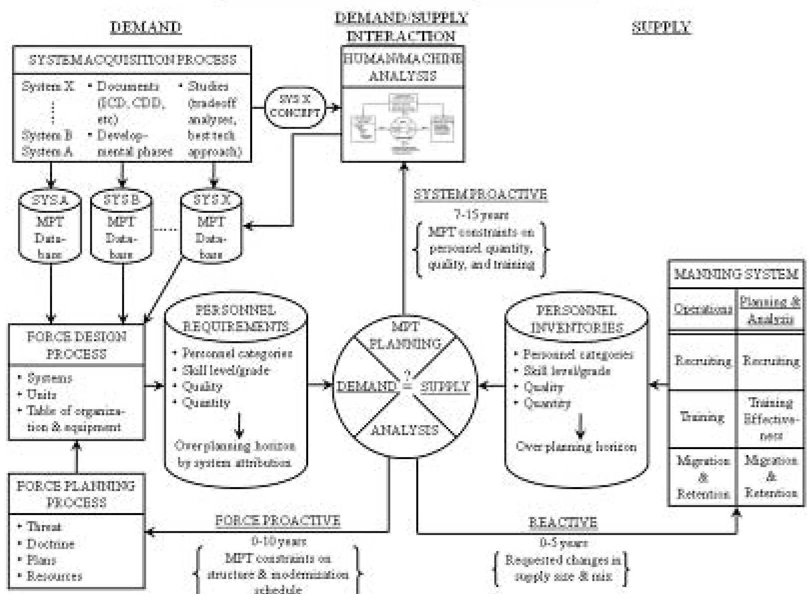
Domain-focused (stove-piped) model (1960s)

Booher's engineering-oriented model (1990s)

NPS systems-oriented model (Miller & Shattuck)

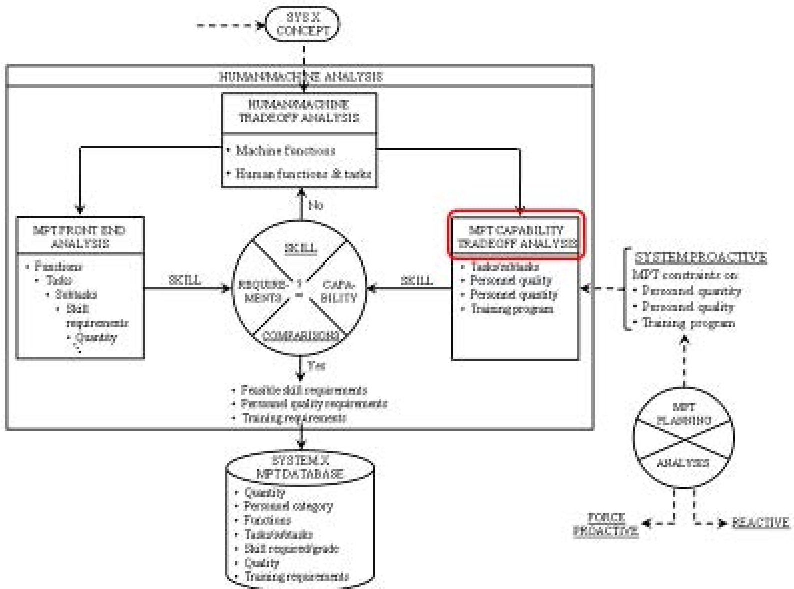


Macro-HSI Perspective:



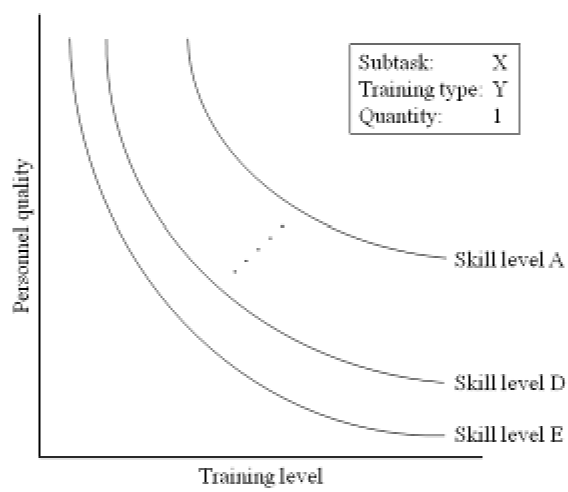
(DePuy & Bonder, 1982)

Micro-HSI Perspective:



(DePuy & Bonder, 1982)

MPT Capability Tradeoff Analysis:



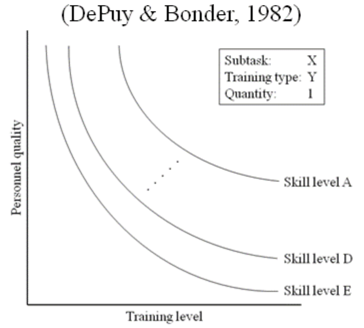
(DePuy & Bonder, 1982)

Formulating an Analytic Approach:

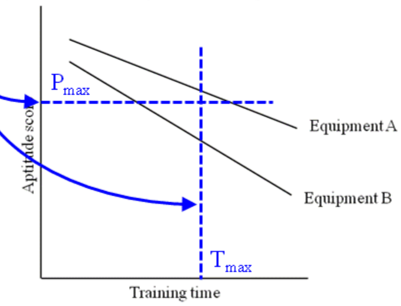
Two MPT (HSI) trade spaces (DePuy & Bonder, 1982)

- Premise: Match MPT supply/demand using reactive & force/system proactive processes
- Macro vs. micro perspectives

MPT Capability Tradeoff Analysis
(DePuy & Bonder, 1982)



Isoperformance
(Jones, Kennedy, & colleagues, 1985-2004)



Isoperformance Methodology:

Def: Quantitative tradeoff methodology based on idea that specified level of performance can be produced by more than one combination of determinants

Approach:

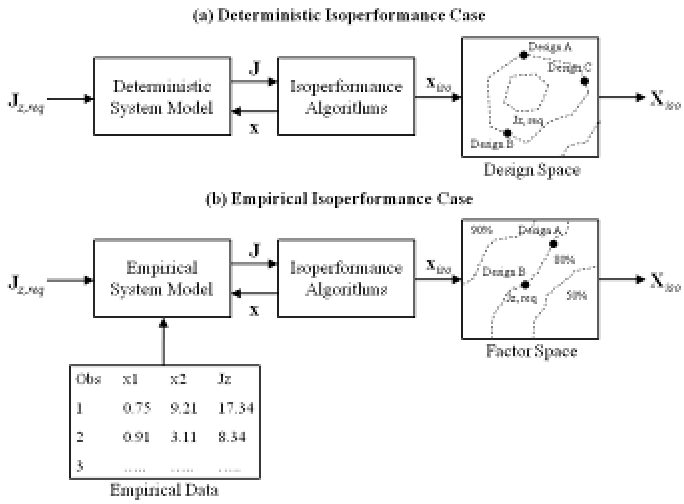
1) Data-analytic procedure based on a model
2) Specify a criterion and level of confidence (assurance level)
3) Derive isoperformance equation
4) Fix performance at criterion / assurance level and solve to identify equivalent sets of determinants

$$\mathrm{E}[R] = b_0 + b_1 A + b_2 B \leftrightarrow \mathrm{E}[R] = r^* + z_\alpha \sigma_r$$

$$r^* = b_0 + b_1 A + b_2 B - z_\alpha \sigma_r$$

(Jones, Kennedy & colleagues, 1985, 1987, 1988, 1989, 1990, 1992, 1993, 1996, 2000, 2004)

Isoperformance Modeling Approaches:



(de Weck & Jones, 2006)

Economically Defining HSI Trade Space:

- Challenge – performance of interest has many determinants (often ≥ 5, 15-30 likely)
- Simonian approach of progressive iteration
  - Fractional factorial study design (2 levels) → screening experiment
  - Pareto chart to define important few (effect size)
  - Advanced experimental designs (3 levels) → non-linear
  - Response surface model
  - Explore trade space → Isoperformance curves

(Simon, 1970, 1973, 1974, 1975, 1976b, 1977, 1978, 1984, 1985, 1987; Simon & Roscoe, 1981)

Coupling Isoperformance with Utility Analysis:

Jones & Kennedy (1990) suggest need but do not develop concept

Again, consider case: $r^* = b_0 + b_1 A + b_2 B - z_\alpha \sigma_r$

Proposition: Instead of fixing r^* (i.e., treating r^* as data)

- Consider A, B, and r^* as decision variables constrained to feasible relationships defined by isoperformance equation
- Use physical programming to determine optimal values in terms of overall utility

Paradigm allows tradeoffs between overall system performance and individual HSI domain considerations (logistics/sustainment)

Conclusions:

Overarching goal of HSI: tradeoffs that are organizationally net positive

Proposed modified Jones-Kennedy-Simon approach:

1) Macro-HSI considerations used to bound/constrain micro-HSI trade space
2) Simonian approach to define HSI drivers & map trade space
3) Use isoperformance to incorporate HSI considerations in systems analyses (tradeoff equations)
4) Couple isoperformance with utility analysis (physical programming)

Problem Statement:

Based on a coherent synthesis of history, we have developed the fundamentals for logically thinking about HSI…

Now, can we apply detailed behavioral and operations research techniques to our HSI planning process?

Three Illustrative Case Studies:

1) Isoreliability Models for HSI Domain Tradeoffs — Choosing a Personnel Supply Source for Future Unmanned Aircraft System Operators

2) HSI Domain Tradeoffs in Non-Technical Systems — Improving Soldier Basic Combat Training

3) HSI Domain Tradeoffs in Optimized Manning — The Task Effectiveness Scheduling Tool (TEST)

Isoreliability Models for HSI Domain Tradeoffs — Choosing a Personnel Supply Source for Future Unmanned Aircraft System Operators

5. Improving Domain Synthesis/Analysis

Case Study 1 — Opportunistic Dataset

Background:

USAF Corona South 4-star general officer summit (1997) tasked AFRL to examine impact of prior flying experience on training for RQ-1 Predator UAS

Schreiber et al (2002) evaluated impact of personnel category on time to train Predator pilot skills & performance on reconnaissance task using UAVSTE (AFRL-HE-AZ-TR-2002-0026)

Study proven surprisingly limited in providing insight into Predator HSI problems

Data:

Original datasets (3) obtained from investigator (D.L.)

- Basic maneuver task
- Landing task
- Reconnaissance task

N = 93 participants (excluding current Predator pilots)

Independent variable: Personnel category (categorical, 6 levels)

Dependent variables (continuous):

- Number of trials to achieve criterion performance (basic maneuver & landing tasks)
- Total time-on-target (reconnaissance tasks)

Reformulation as HSI Problem:

Independent variables:

- Personnel category (categorical)
- Training (continuous)

Dependent variable: Proficient (dichotomous)

Study questions:

1) Quantitatively assess the relative importance of personnel and training domains and their interaction on task proficiency?
2) Adapt isoperformance methodology to consider personnel and training domain tradeoffs in terms of task proficiency?
3) Aggregate isoperformance models across system functions?

Isoperformance:

Def: Quantitative tradeoff methodology based on idea that specified level of performance can be produced by more than one combination of determinants (Jones, Kennedy & colleagues)

Objective: Redact tradeoffs in terms of *isoreliability*

Data analysis steps:

1) Data-analytic procedure based on a model
2) Specify a criterion and level of confidence (assurance level)
3) Derive isoreliability equation
4) Fix reliability at criterion / assurance level and solve to identify equivalent sets of determinants

Accommodate categorical determinant?

Step 1 – Data Analytic Procedure:

Dependent variable, y_j, takes on only two possible values, 0 and 1, depending on whether the j^{th} participant isn't or is proficient respectively

A reasonable probability model for y_j is the binomial with $P(y_j = 1) = \pi_j$:

$$\log\left(\frac{\hat{y}}{1-\hat{y}}\right) = \beta_0 + \sum_{i=1}^{6} \beta_i x_{ij} + \sum_{i=2}^{6} \beta_{1i} x_{1j} x_{ij}$$

where:

x_1 = Trials $[0, +\infty)$
x_2 = Civil instrument pilots $\{0,1\}$
x_3 = Civil private pilots $\{0,1\}$
x_4 = Predator selectees $\{0,1\}$
x_5 = T-1 graduates $\{0,1\}$
x_6 = T-38 graduates $\{0,1\}$

Fitted model for landing task data:

$$\log\left(\frac{\hat{y}}{1-\hat{y}}\right) = -3.5976 + 0.0325x_1 + 0.8351x_2 + 1.8278x_3 - 0.4717x_4 + 1.5178x_5$$
$$+ 0.7784x_6 + 0.0317x_1x_2 - 0.0025x_1x_3 + 0.0531x_1x_4 - 0.0054x_1x_5 + 0.0200x_1x_6$$

Steps 2 & 3 – Deriving Isoreliability Equation:

The analytic model for y_j:

$$y_j = \frac{\exp\left(\beta_0 + \sum_{i=1}^{6} \beta_i x_{ij} + \sum_{i=2}^{6} \beta_{1i} x_{1j} x_{ij}\right)}{1 + \exp\left(\beta_0 + \sum_{i=1}^{6} \beta_i x_{ij} + \sum_{i=2}^{6} \beta_{1i} x_{1j} x_{ij}\right)} + \varepsilon_j$$

Fitting the model to data:

$$E\left(y_j\right) = \pi_j = \frac{\exp\left(\hat{\beta}_0 + \sum_{i=1}^{6} \hat{\beta}_i x_{ij} + \sum_{i=2}^{6} \hat{\beta}_{1i} x_{1j} x_{ij}\right)}{1 + \exp\left(\hat{\beta}_0 + \sum_{i=1}^{6} \hat{\beta}_i x_{ij} + \sum_{i=2}^{6} \hat{\beta}_{1i} x_{1j} x_{ij}\right)} \qquad (1)$$

Determine what π_j should be so that the j^{th} participant achieves a specified probability with a desired assurance level, α:

$$E\left(y_j\right) = \pi_j = \frac{\exp\left(\log\left(\frac{\pi_{spec}}{1-\pi_{spec}}\right) + z_\alpha \sqrt{\mathrm{Var}\left(\hat{\beta}_0 + \sum_{i=1}^{6} \hat{\beta}_i x_{ij} + \sum_{i=2}^{6} \hat{\beta}_{1i} x_{1j} x_{ij}\right)}\right)}{1 + \exp\left(\log\left(\frac{\pi_{spec}}{1-\pi_{spec}}\right) + z_\alpha \sqrt{\mathrm{Var}\left(\hat{\beta}_0 + \sum_{i=1}^{6} \hat{\beta}_i x_{ij} + \sum_{i=2}^{6} \hat{\beta}_{1i} x_{1j} x_{ij}\right)}\right)} \qquad (2)$$

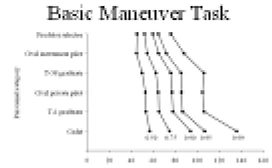
Combining Equations 1 and 2, rearranging terms, and using matrix notation:

$$\log\left(\frac{\pi_{spec}}{1-\pi_{spec}}\right) = \mathbf{x}_j'\hat{\boldsymbol{\beta}} - z_\alpha \sqrt{\mathrm{Var}\left(\mathbf{x}_j'\hat{\boldsymbol{\beta}}\right)}$$

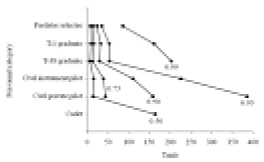
where $\mathrm{Var}\left(\mathbf{x}_j'\hat{\boldsymbol{\beta}}\right) = \mathbf{x}_j'\left(\mathbf{X}'\mathbf{V}\mathbf{X}\right)^{-1}\mathbf{x}_j$ and $\mathbf{X}'\mathbf{V}\mathbf{X}$ is the covariance matrix of model parameters

Steps 4 – Identify Equivalent Sets of Determinants:

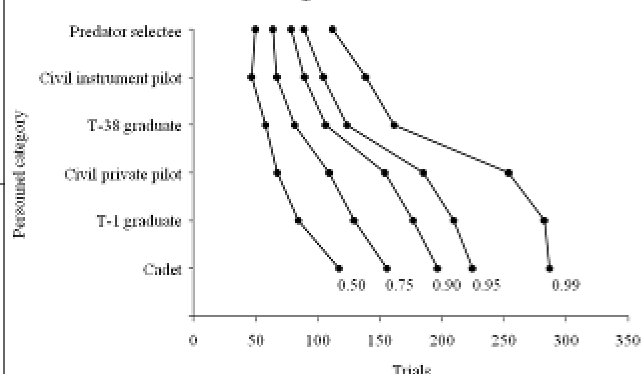
Basic Maneuver Task



Reconnaissance Task



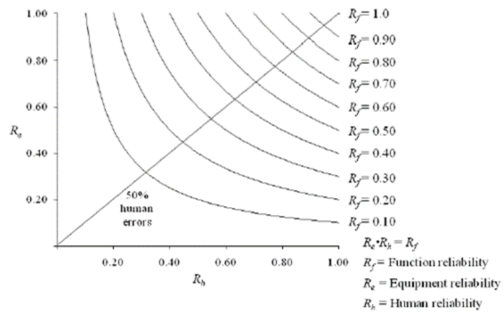
Landing Task



Further Insights:

Nelson, Schmitz, and Promisel (1984) defined reliability of a system function, *f*, as follows:

$$R_f = R_e \cdot R_h$$



Just showed that *human reliability* can be expressed in terms of training time and personnel category:

$$R_h(x_1, x_2) = \frac{1}{1+\exp\left(-\boldsymbol{x}\hat{\boldsymbol{\beta}} + z_\alpha\sqrt{\mathrm{Var}\left(\boldsymbol{x}\hat{\boldsymbol{\beta}}\right)}\right)}$$

The overall reliability (or the probability of successful performance) of a system function, *f*, can be defined as follows:

$$R_f = R_{e,f} \cdot R_{h,f}(x_1, x_2)$$

Now can avail ourselves of basic system models to aggregate results into a system level estimate

It is also possible to define the following relationship:

$$R_h(x_1, x_2) = 1 - F_h(x_1, x_2)$$

where $F_h(x_1, x_2)$ is conditional probability operator will fail (*human unreliability function*).

Once decide on personnel selection and training policy, (x_1^*, x_2^*), we have fixed probability of success, $p = R(x_1^*, x_2^*)$, and fixed probability of failure, $q = F(x_1^*, x_2^*)$.

Geometric distribution commonly used to model the number of cycles to failure for items that have a fixed probability of failure associated with each cycle.

If system cycle lengths, C_i, are independent and identically distributed random variables with an expected cycle length of $\mathrm{E}[C]$, then model for time until first failure, *T*:

$$T = \sum_{i=1}^{N} C_i$$

Expected time until first human failure, $\mathrm{E}[T]$, computed as follows:

$$\mathrm{E}[T] = \mathrm{E}[N]\mathrm{E}[C] = \frac{1}{q}\mathrm{E}[C] = \frac{1+\exp\left(-\boldsymbol{x}^*\hat{\boldsymbol{\beta}} + z_\alpha\sqrt{\mathrm{Var}\left(\boldsymbol{x}^*\hat{\boldsymbol{\beta}}\right)}\right)}{\exp\left(-\boldsymbol{x}^*\hat{\boldsymbol{\beta}} + z_\alpha\sqrt{\mathrm{Var}\left(\boldsymbol{x}^*\hat{\boldsymbol{\beta}}\right)}\right)}\mathrm{E}[C]$$

Expected frequency of system operator failures, $\mathrm{E}[Y]$, or human failure rate:

$$\mathrm{E}[Y] = \frac{1}{\mathrm{E}[T]} = \frac{\exp\left(-\boldsymbol{x}^*\hat{\boldsymbol{\beta}} + z_\alpha\sqrt{\mathrm{Var}\left(\boldsymbol{x}^*\hat{\boldsymbol{\beta}}\right)}\right)}{\left(1+\exp\left(-\boldsymbol{x}^*\hat{\boldsymbol{\beta}} + z_\alpha\sqrt{\mathrm{Var}\left(\boldsymbol{x}^*\hat{\boldsymbol{\beta}}\right)}\right)\right)\mathrm{E}[C]}$$

Given severity rating, *s*, for seriousness of effects or impact of a system operator's failure to satisfactorily perform a function or task, can calculate a risk assessment value (RAV):

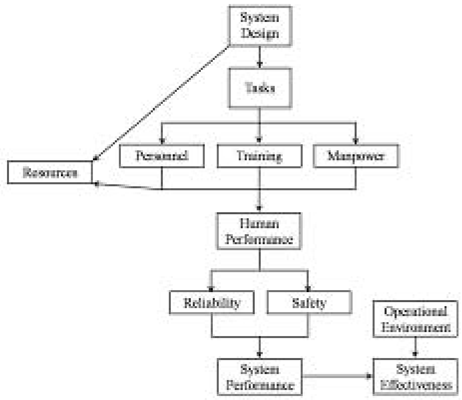
$$\mathrm{RAV} = \mathrm{E}[Y] \cdot s$$

Implications:

1) Safety domain conceptualized as function of HFE, personnel, and training domains.

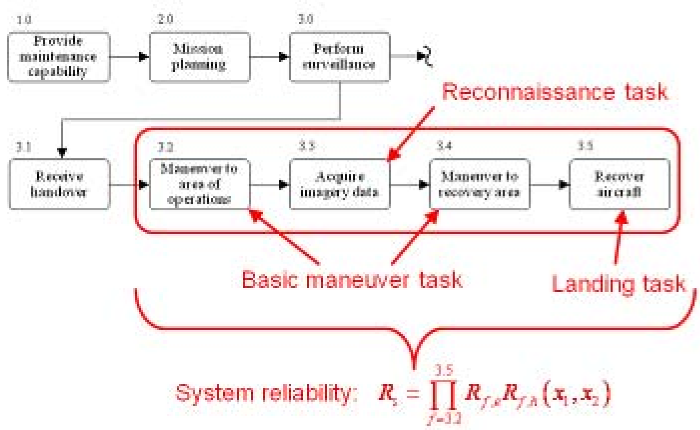
2) Safety probabilistically related to presence of satisfactory performance, which can be expressed in terms of human reliability

3) Hierarchical relationship of HSI domains



Aggregate Isoreliability Analysis:

Reliability block diagram (Nagy, Kalita, & Eaton, 2006)



Mathematical Program:

Indices

$f \triangleq$ function ($f = 3.2, 3.3, 3.4, 3.5$)

$p \triangleq$ personnel category (p = Predator selectee, …, cadet)

$t \triangleq$ task (t = maneuver, landing, recon)

Data

$e_f \triangleq$ the equipment reliability for function f

$m_p \triangleq$ the personnel costs for an operator from personnel category p

$c_t \triangleq$ the hourly cost for training task t

$l \triangleq$ the average duration of a landing trail (in minutes)

$\underline{r} \triangleq$ the lower limit on acceptable total system reliability

$\overline{t} \triangleq$ the upper limit on available training time (in minutes)

Decision variables (non-negative or binary)

$x_{1,t} \triangleq$ the amount of training provided for task t

$$x_{2,p} \triangleq \begin{cases} 1 \text{ if an operator from personnel category } p \text{ is selected} \\ 0 \text{ otherwise} \end{cases}$$

$h_{f,p} \triangleq$ carrier variable for the human reliability value for function f given personnel category p

$w_f \triangleq$ carrier variable for amount of training provided for function f

$y_t \triangleq$ carrier variable for training time (in hours) for task t

Formulation

$$\max \quad \frac{\prod_f \left[e_f \left(\sum_p h_{f,p} x_{2,p} \right) \right]}{\sum_p m_p x_{2,p} + \sum_t c_t y_t}$$

s.t.

$$w_{3.2} = w_{3.4} = x_{1,\text{maneuver}}$$

$$w_{3.3} = x_{1,\text{recon}}$$

$$w_{3.5} = x_{1,\text{landing}}$$

$$h_{f,p} = g_{f,p}(w_f) \quad \forall f, p$$

$$\sum_p x_{2,p} = 1$$

$$\underline{r} \le \prod_f \left[e_f \left(\sum_p h_{f,p} x_{2,p} \right) \right]$$

$$y_{\text{maneuver}} = \frac{x_{1,\text{maneuver}}}{60}$$

$$y_{\text{landing}} = \frac{l x_{1,\text{landing}}}{60}$$

$$y_{\text{recon}} = \frac{10 x_{1,\text{recon}}}{60}$$

$$\overline{t} \ge \sum_t y_t$$

Cost Estimating Data:

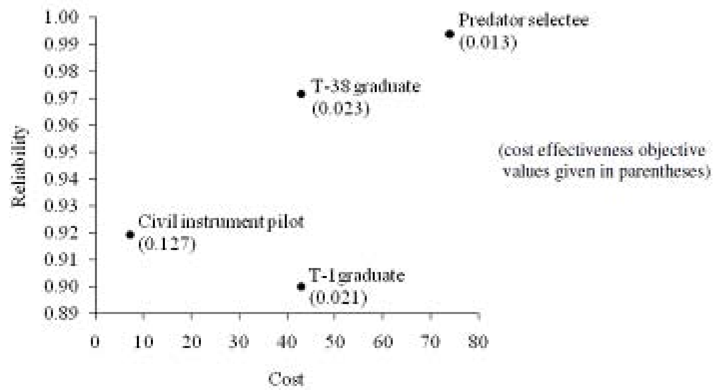
Personnel category	Manpower cost elements	Estimates (FY05)	Normalized costs[8]
Predator selectee	SMCR[1] O-3	$100,833[5]	73.783
	SUPT[2]	392,861[7]	
	B-52 IQT[3]	292,190[7]	
	Total	$785,934	
T-38 graduate	SMCR[1] O-1	$ 62,982[5]	42.794
	SUPT[2]	392,861[7]	
	Total	$455,843	
T-1 graduate	SMCR[1] O-1	$ 62,982[5]	42.794
	SUPT[2]	392,861[7]	
	Total	$455,843	
Civil instrument pilot	SMCR[1] O-1	$ 62,982[5]	7.039
	IFT[4]	5,500[7]	
	Instrument rating	6,500[7]	
	Total	$74,982	
Civil private pilot	SMCR[1] O-1	$ 62,982[5]	6.429
	IFT[4]	5,500[7]	
	Total	$68,482	
Cadet	SMCR[1] Cadet	$10,652[6]	1.000

[1]SMCR = Standard military compensation rate
[2]SUPT = Specialized undergraduate pilot training
[3]IQT = Initial qualification training
[4]IFT = Initial flight training
[5]Source: Dahlman, 2005
[6]Source: DoD Comptroller, 2010
[7]Source: Hoffman & Kamps, 2005
[8]Relative to cadet

Results:

The optimal solution has the following applicable non-zero variables:

$x^*_{1M} = 153.4$	$w^*_{3,2} = 153.4$	$y^*_M = 2.6$	$h^*_{3,2,1} = 1.0000$
$x^*_{1L} = 163.1$	$w^*_{3,3} = 143.1$	$y^*_L = 13.6$	$h^*_{3,3,1} = 0.9220$
$x^*_{1R} = 143.1$	$w^*_{3,4} = 153.4$	$y^*_R = 23.8$	$h^*_{3,4,1} = 1.0000$
$x^*_{2,1} = 1$	$w^*_{3,5} = 163.1$		$h^*_{3,5,1} = 0.9970$



Conclusions:

Fulfilled Weisz's HSI paradigm → took experiment from behavioral sciences and transferred results into mathematical models tractable to optimization techniques of OR

Demonstrated feasibility of including logical decision variables in isoperformance models → incorporated these isoperformance models into discrete optimization models to analyze aggregated functions (systems analysis)

Applied isoperformance methodology to construct of human reliability

- Advantage over THERP in trade-off studies
- Provides construct for analyzing data generated by Siegel/Wolf models (IMPRINT)

HSI Domain Tradeoffs in Non-Technical Systems — Improving Soldier Basic Combat Training

6. Improving Domain Synthesis/Analysis

Case Study 2 — Prospective Dataset

Background:

Sleep deprivation prevalent in military training & education programs (Killgore et al., 2008; Miller, 2005; Miller et al., 2008)

Military recruits adolescents or young adults with distinct, biologically-driven sleep-wake patterns (Carskadon et al., 1997, 1998; Wolfson & Carskadon, 2003)

- Delayed bedtimes, later awakenings & longer sleep periods
- May require 8.5–9.25 hrs sleep per night for optimal performance

Multiple nights of less than 8 hrs sleep → sleep debt & fatigue, the effects of which include:

- Decreased vigilance, adverse mood changes, perceptual & cognitive decrements (Krueger, 1990; Belenky et al., 2003; van Dongen et al., 2003)
- Impaired judgment & increased risk taking (Killgore, Balkin, & Wesensten, 2006)
- Decreased marksmanship (Tharion, Shunkitt-Hale, & Lieberman, 2003; McLellan et al., 2005)

Background

Motivation only partially compensates for effects of sleep deprivation (Pigeau, Angun, O'Neil, 1995)

Ability of individuals to learn & retain information reduced by sleep deprivation

- Learning curves drop for adolescents with 4–6 vs. 8 hrs sleep (Graham, 2000)
- Navy recruit academic performance improved with change in sleep regimen from 6 to 8 hrs (Andrews, 2004)
- Positive correlation between soldier test scores & daily sleep (Killgore et al., 2008)

Correlations between sleep / fatigue & safety / health (Moldofsky, 1995; Lange et al., 2003; Thorne et al., 1992)

Study Hypotheses:

H1: Participants on the modified, phase-delayed sleep schedule will obtain more daily sleep than participants following the standard BCT schedule

H2: Participants on the modified sleep schedule will have less decrement in mood state than participants following the standard BCT sleep schedule

H3: Participants on the modified sleep schedule will exhibit greater improvement in basic rifle marksmanship scores than participants following the standard BCT sleep schedule

H4: Participants on the modified sleep schedule will exhibit greater improvement in physical fitness scores than participants following the standard BCT sleep schedule

Study Hypotheses:

H5: The odds of participants on the modified sleep schedule reporting occupationally significant fatigue will be lower than that for participants following the standard BCT sleep schedule

H6: The odds of participants on the modified sleep schedule reporting poor sleep quality will be lower than that for participants following the standard BCT sleep schedule

H7: The odds of participants on the modified sleep schedule attriting from training will be lower than that for participants following the standard BCT sleep schedule

Methods:

Study design: Quasi-experimental

Inclusion criteria: Soldier assigned to C or B/3-10 IN BN, FLW, starting BCT on 14 / 21 Aug10

B: $N - O_1 - X - O_2 - O_3 - O_4 - O_5 - O_6 - O_7 - O_8 - O_9$

C: $N - O_1 - O_2$ —— $O_3 - O_4 - O_5 - O_6 - O_7 - O_8 - O_9$

Non-random assignment to company (N)

Random assignment of company to treatment condition (X)

Treatment condition (X) = modified, phase delayed sleep schedule (2300–0700)

Comparison condition = standard BCT sleep schedule (2030–0430)

Data Collection Instruments & Variables:

↓Data Event Week→	1	2	3	4	5	6	7	8	9
Actigraphy*	X	X	X	X	X	X	X	X	X
Army Physical Fitness Test			X			X		X	
Basic Rifle Marksmanship					X				
Epworth Sleepiness Scale	X								X
Morningness-Eveningness Questionnaire	X								
NEO Five-Factor Inventory	X								
Pittsburgh Sleep Quality Index	X								X
Profile of Mood States	X	X	X	X	X	X	X	X	X
Response to Stressful Experiences Scale	X								
Study Questionnaires	X								X

*Actigraphy data collected on a random sample of study participants

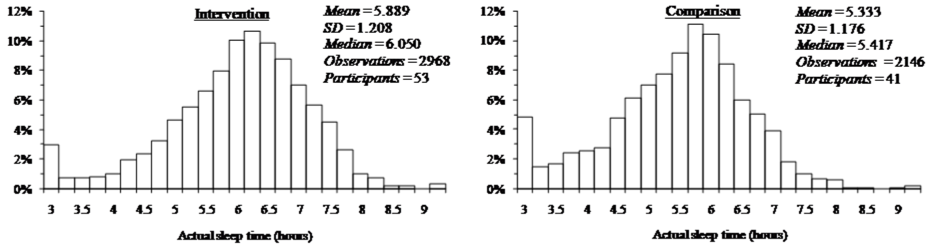
Participants:

Variable	Intervention	Comparison	*p*-value
N	209	183	
Actigraphy	53 (25%)	41 (22%)	
BMI, median (IQR)	25.4 (22.9–28.4)	23.6 (21.6–26.8)	0.021
Component			
National Guard	72 (34.4%)	58 (31.7%)	
Regular	82 (39.2%)	109 (59.6)	< 0.001
Reserves	55 (26.3%)	16 (8.7)	
NEO-FFI, median (IQR)			
Neuroticism	52 (45–59)	55 (47–63)	0.012
Conscientiousness	50 (43–57)	46 (38–53)	0.003
Pittsburgh Sleep Quality Index			
Global score, median (IQR)	6 (4–9)	7 (5–10)	0.048
Poor sleep quality (score > 5)	123 (59%)	129 (71%)	0.016
Response to Stressful Experiences Scale, median (IQR)	69 (60–78)	67 (57–75)	0.008

Results:

H1: Participants on the modified, phase-delayed sleep schedule will obtain more daily sleep than participants following the standard BCT schedule

Supported



Intervention group 33 minutes more sleep than comparison group ($p < 0.001$)

OR episode of daily sleep less than NSF recommendation 3.8 (3.0–4.8) for comparison vs. intervention groups

Results:

H2: Participants on the modified sleep schedule will have less decrement in mood state than participants following the standard BCT sleep schedule

Weakly Supported

Over course of BCT, general trend for all participants to report decreased:
- Tension–anxiety
- Depression–dejection
- Fatigue–inertia
- Confusion–bewilderment

Intervention group: less anger–hostility & lower total mood disturbance (TMD) scores early in training → differences diminish over time

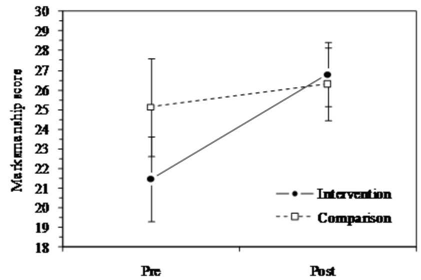
Intervention group: greater feelings of vigor (modest effect size)

Effects of chronotype: mixed results overall, intervention group evening–types better mood

Results:

H3: Participants on the modified sleep schedule will exhibit <u>greater improvement in basic rifle marksmanship scores</u> than participants following the standard BCT sleep schedule

Supported (actigraphy subsample)



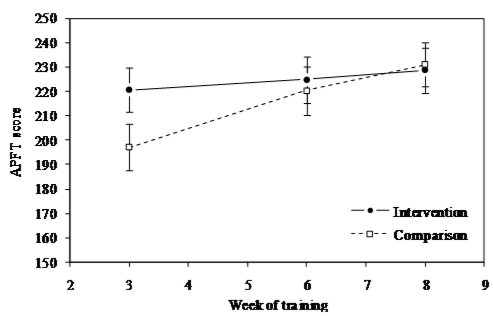
Significant interaction effect (p = 0.017, eta^2 = 0.071)

Significant effect for week $t^* - 1$ average sleep (p = 0.047, eta^2 = 0.050)

Results:

H4: Participants on the modified sleep schedule will exhibit <u>greater improvement in physical fitness scores</u> than participants following the standard BCT sleep schedule

Not Supported



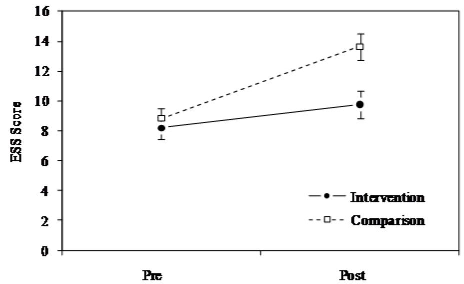
Significant interaction effect (p = 0.001, eta^2 = 0.017)

No significant effect for weekly average sleep (actigraphy subsample)

Results:

H5: The odds of participants on the modified sleep schedule reporting occupationally significant fatigue will be lower than that for participants following the standard BCT sleep schedule

Supported



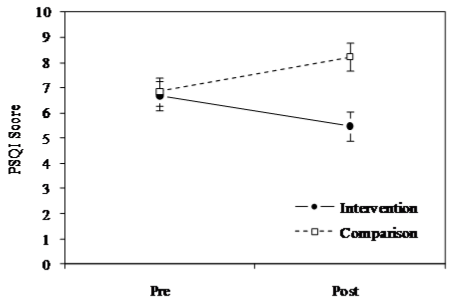
Significant interaction effect ($p < 0.001$, $eta^2 = 0.060$)

OR participant with occupationally significant fatigue (ESS score >10) in comparison vs. intervention group: pre = 1.2 (0.7–1.9); post = 2.3 (1.5–3.7)

Results:

H6: The odds of participants on the modified sleep schedule reporting poor sleep quality will be lower than that for participants following the standard BCT sleep schedule

Supported



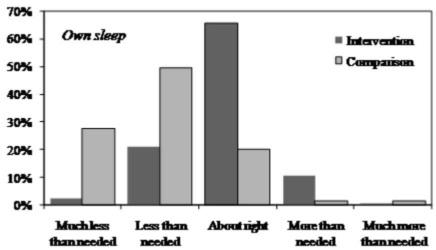
Significant interaction effect ($p < 0.001$, $eta^2 = 0.075$)

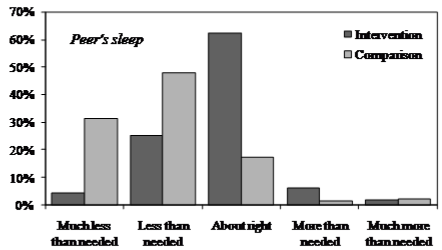
OR participant with poor sleep quality (PSQI > 5) in comparison vs. intervention group: pre = 1.7 (1.1–2.6); post = 5.5 (3.3–9.0)

Results:

H6 (cont):

Ordinal sleep ratings:





Mean rank higher (better sleep) for intervention group (U = 5164.5, $p < 0.001$)

Results:

H7: The odds of participants on the modified sleep schedule attriting from training will be lower than that for participants following the standard BCT sleep schedule

Not Supported

Attrites

Intervention	Comparison
35 (16.7%)	33 (18.1%)

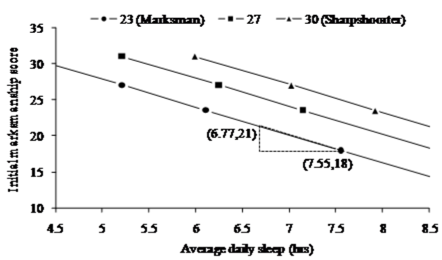
$$\chi^2 = 0.130, p = 0.718$$

Fitted binary logistic regression model: BMI, NEO-FFI neuroticism, POMS D-factor, & sex

HSI Analyses:

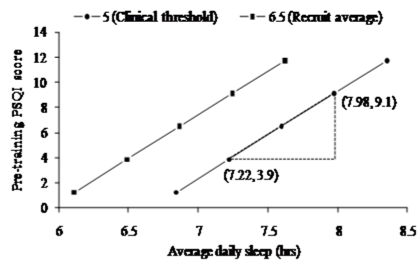
Weisz's HSI paradigm: behavioral sciences → tradeoff functions (OR/systems analysis)

Basic Rifle Marksmanship Isoperformance Model



Sleep improves marksmanship:
16 min ≈ sleep 1 pt (score)

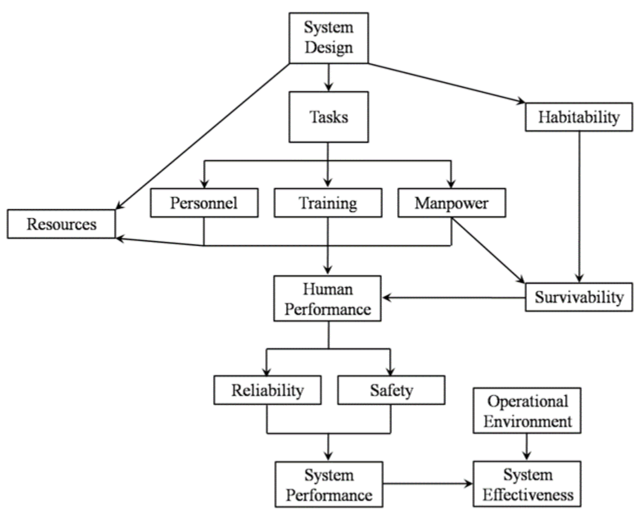
Sleep Quality Isoperformance Model



Sleep improves health:
9 min sleep ≈ 1 pt (score)

Implications:

Extend basic systems integration model for HSI:



Conclusions:

New paradigm for considering training effectiveness: focus on hours spent sleeping rather than activities during wake periods

Accommodating phase delay in adolescent circadian cycle → increased total daily sleep & modest improvements in indicators of daytime functioning

HSI tradeoff analyses provide empirical foundation to quantitatively assess contribution of sleep to Soldier well-being & performance (HSI application to non-technical system)

Quantity & quality of sleep limited are resource variables to be considered as part of human factors contribution to systems analyses

HSI Domain Tradeoffs in Optimized Manning — The Task Effectiveness Scheduling Tool (TEST)

7. Improving Domain Synthesis/Analysis

Case Study 3 — Modeling & Simulation

Background:

Mathematical models of sleep and circadian process in existence for more than 2 decades

Applied biomathematical models use info about sleep history, duration of wakefulness & circadian phase to predict performance capability & risk (Neri, 2004)

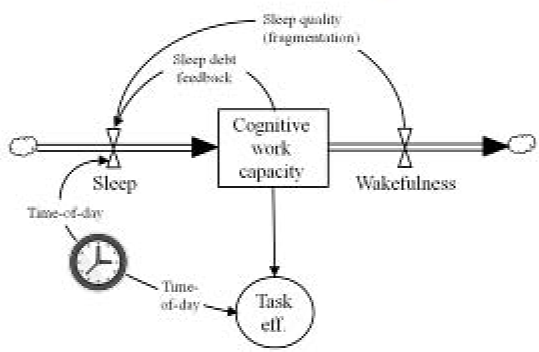
DoD developed Sleep, Activity, Fatigue, and Task Effectiveness (SAFTE) model, implemented in Fatigue Avoidance Scheduling Tool (FAST) (Hursh et al., 2004)

SAFTE Model:

Independent variables:

- Schedule (manpower & survivability domains)
- Sleep environment (habitability domain)

Dependent variable: Task effectiveness (performance)



Problem Statement:

FAST paradigm: given schedule → forecast task effectiveness

What about inverse questions?

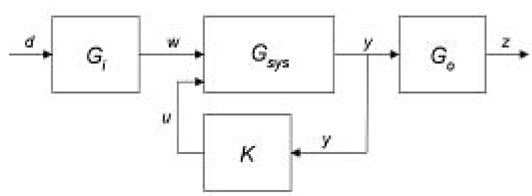
Task effectiveness threshold value → Optimal schedule in terms of timing of sleep-wake periods & assignment of performance-sensitive duties

Task effectiveness threshold value → Minimum number of personnel

Systematic exploration of solution space

Study Questions:

Assume a generic dynamic system with system controller, K,



Then:

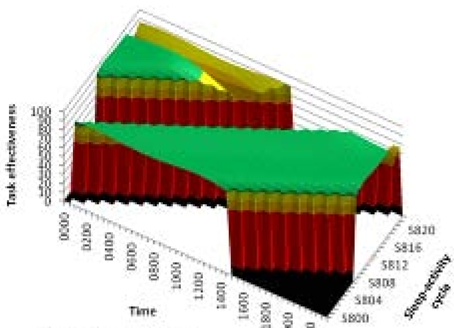
1) Given an *a priori* task effectiveness requirement, what is the minimum number of individuals needed to staff function K?
2) Given this minimum number, how should duty periods be scheduled to maximize average task effectiveness?

Simulations:

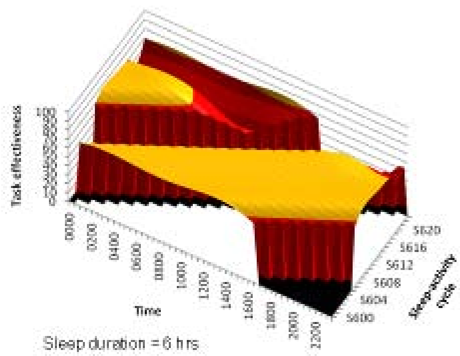
Schedule (72-levels) X Time Period (48) X Sleep Quality (4)

Each schedule and sleep quality condition simulated over 30-day period

Task effectiveness set to 0% ±1 hr from as well as during sleep periods



Sleep duration = 8 hrs
Sleep quality = good



Sleep duration = 6 hrs
Sleep quality = fair

Mathematical Program:

Indices and [Cardinality]

$q \in Q$ — set of ordinal ratings of sleep quality [~4].

$s \in S$ — set of wake-sleep schedules [~72].

$t \in T$ — set of time periods [~48].

Data and [Units]

req_eff — required human task effectiveness [%]

$safte\_data^{q}_{s,t}$ — predicted task effectiveness for time period t when following schedule s with sleep quality q [%]

work_rule — organizational limit on maximum hours of service [periods]

Variables (non-negative or binary)

$ASSIGN_{s,t}$ — binary decision variable to assign a person following schedule s to cover time period t

$D_{s,t}$ — difference variable used to determine a change in the state (i.e., on or off duty) of a person following schedule s at time period t

$MANPOWER_s$ — binary decision variable to utilize a person on schedule s

Constraints

(C1) $\sum_{s} ASSIGN_{s,t} = 1 \ \forall\ t$

(C2) $\sum_{t} ASSIGN_{s,t} \le work\_rule \ \forall\ s$

(C3) $\sum_{s} safte\_data_{s,t} ASSIGN_{s,t} \ge req\_eff \ \forall\ t$

(C4) $D_{s,t} \ge ASSIGN_{s,t} - ASSIGN_{s,t-1} \ \forall\ s,t > 1$

(C5) $D_{s,t} \ge -ASSIGN_{s,t} + ASSIGN_{s,t-1} \ \forall\ s,t > 1$

(C6) $\sum_{t|t>1} D_{s,t} \le 2 \ \forall\ s$

(C7) $MANPOWER_s \ge ASSIGN_{s,t} \ \forall\ t,s$

(C8) $ASSIGN_{s,t} \in \{0,1\} \ \forall\ t,s$

(C9) $MANPOWER_s \in \{0,1\} \ \forall\ s$

(C10) $0 \le D_{s,t} \le 1 \ \forall\ s,t > 1$

Objective

$$\text{Minimize } Z = \sum_s MANPOWER_s$$

Once the value of the manpower objective is minimized (that is, Z* is determined), a new constraint is created

$$\text{(C11) } Z^* \geq \sum_s MANPOWER_s$$

The program is then solved for the following objective:

$$\text{Maximize } \frac{\sum_s \sum_t safte\_data_{s,t} ASSIGN_{s,t}}{48}$$

Scenario 1a – High Task Effectiveness Criterion:

X Assigned for duty | Available for other duty | Unavailable | Average Task Effectiveness = [illegible]

	0000	0030	0100	0130	0200	0230	0300	0330	0400	0430	0500	0530	0600	0630	0700	0730	0800	0830	0900	0930	1000	1030	1100	1130
1																								X
2																								
3												X	X	X	X									
4																X	X	X	X					
5																				X	X	X	X	
6																								
7	X	X	X	X																				
8																								
9																								
10					X	X	X	X	X	X	X													

	1200	1230	1300	1330	1400	1430	1500	1530	1600	1630	1700	1730	1800	1830	1900	1930	2000	2030	2100	2130	2200	2230	2300	2330
1	X	X																						
2			X	X	X	X																		
3																								
4																								
5																								
6																			X	X	X	X	X	
7																								X
8							X	X	X	X	X	X												
9													X	X	X	X	X	X						
10																								

Task effectiveness criterion = 95%

Sleep quality = good

Scenario 1b – High Task Effectiveness Criterion:

X Assigned for duty   Available for other duty   Unavailable   Average Task Effectiveness = 94.69%

	0000	0030	0100	0130	0200	0230	0300	0330	0400	0430	0500	0530	0600	0630	0700	0730	0800	0830	0900	0930	1000	1030	1100	1130
1													X	X	X	X	X	X	X	X	X	X	X	X
2	X	X	X	X	X	X	X	X	X	X	X	X												

	1200	1230	1300	1330	1400	1430	1500	1530	1600	1630	1700	1730	1800	1830	1900	1930	2000	2030	2100	2130	2200	2230	2300	2330
1	X	X	X	X	X	X	X	X	X	X	X	X												
2													X	X	X	X	X	X	X	X	X	X	X	X

Task effectiveness criterion = 90%
Sleep quality = good

Scenario 2 – Organizational Hours-of-Work Rules:

X Assigned for duty   Available for other duty   Unavailable   Average Task Effectiveness = 94.64%

	0000	0030	0100	0130	0200	0230	0300	0330	0400	0430	0500	0530	0600	0630	0700	0730	0800	0830	0900	0930	1000	1030	1100	1130
1													X	X	X	X	X	X	X	X	X	X	X	X
2																								
3	X	X	X	X	X	X	X	X	X	X	X	X												

	1200	1230	1300	1330	1400	1430	1500	1530	1600	1630	1700	1730	1800	1830	1900	1930	2000	2030	2100	2130	2200	2230	2300	2330
1	X	X	X	X	X	X	X	X																
2									X	X	X	X	X	X	X	X								
3																	X	X	X	X	X	X	X	X

Task effectiveness criterion = 90%
Sleep quality = good
10 hours-of-work rule enforced

Scenario 3 – Sleep Quality:

X Assigned for duty   Available for other duty   Unavailable   Average Task Effectiveness = 78.96%

	0000	0030	0100	0130	0200	0230	0300	0330	0400	0430	0500	0530	0600	0630	0700	0730	0800	0830	0900	0930	1000	1030	1100	1130
1																								
2													X	X	X	X								
3																								
4																	X	X	X	X				
5																					X	X	X	X
6																								
7	X	X	X	X																				
8					X	X	X	X	X	X	X	X												

	1200	1230	1300	1330	1400	1430	1500	1530	1600	1630	1700	1730	1800	1830	1900	1930	2000	2030	2100	2130	2200	2230	2300	2330
1	X	X	X	X	X	X	X																	
2																								
3								X	X	X	X	X	X	X										
4																								
5																								
6															X	X	X	X	X	X	X			
7																						X	X	X
8																								

Task effectiveness criterion = 77.5% (FAST criterion line)
Sleep quality = poor

Conclusions:

Systemic and systematic approach to designing staffing and shift scheduling solutions

Systematic in that:

- Uses data from FAST to answer questions of optimality using deterministic process
- Makes explicit boundaries of human capacity

Systemic in that:

- Explores trade space between manpower, survivability, habitability, and human factors engineering domains of HSI
- Facilitates incorporation of HSI considerations in systems analyses

Lessons From Discourse:

Surprisingly—or perhaps not—almost nothing has been written since 1990 that deals with the theory and practice of HSI at the operational level

Primary accomplishments of this discourse:

- Extract lessons learned from historical analysis of emergence of HSI as both philosophy and program
- Apply lessons to develop & illustrate approach to addressing HSI considerations early in WSAP

Discourse appears on face to provide sensible accounting of HSI vis-à-vis pre-MS A activities required by WSARA (2009)

New HSI:

Step 1	Establish SOI* objectives and requirements by reference to containing system(s)
Step 2	Identify containing system(s)' strategic human resources objectives
Step 3	Identify sibling systems (vis-à-vis shared human resources) and their interactions that will be perturbed by the SOI
Step 4	Develop SOI design trade space to complement sibling systems in contributing to containing system(s)' objectives
Step 5	Functionally partition SOI and describe required (emergent) human-system performance in terms of response surfaces that are functions of the domains of HSI
Step 6	Reduce response surfaces to isoperformance (tradeoff) equations for incorporation in system analyses
Step 7	Seek a balanced design (joint optimization) that satisfices SOI objectives and requirements
Step 8	Continuously reassess and rebalance the design throughout the life of the SOI

*SOI = System-of-interest

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
Ft. Belvoir, Virginia

2. Dudley Knox Library
Naval Postgraduate School
Monterey, California

3. Dr. Nita Shattuck
Department of Operations Research
Naval Postgraduate School
Monterey, California

4. Dr. Lawrence Shattuck
Department of Operations Research
Naval Postgraduate School
Monterey, California

5. Dr. Michael McCauley
MOVES Institute
Naval Postgraduate School
Monterey, California

6. Dr. Clifford Whitcomb
Chairman, Department of Systems Engineering
Naval Postgraduate School
Monterey, California

7. Dr. Eugene Paulo
Department of Systems Engineering
Naval Postgraduate School
Monterey, California

8. Col Lex Brown, USAF
Directorate of Human Performance Integration
Brooks City-Base, Texas

9. Col Larry Kimm, USAF
Air Force Human Systems Integration Office
Falls Church, Virginia

10. Dr. Michael Drillings
Director of MANPRINT, Army G-1
Washington, DC

11. Major General Gregg F. Martin, USA
Commandant, U.S. Army War College
Carlisle Barracks, Pennsylvania

12. Mr. Joseph Driskill
Leonard Wood Institute
Fort Leonard Wood, Missouri